

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comogen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 08:46:10 ; Search time 59 seconds

[without alignments]
1431.893 Million cell updates/sec

Title: US-10-007-521-12

Perfect score: 1639

Sequence: 1 MESTVRLTLLAALPLVNS.....TTCVSGTCCXNDYSGQL 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: Genesegp29Jan04:*
2: genesegp1980s:*
3: genesegp2000s:*
4: genesegp2001s:*
5: genesegp2002s:*
6: genesegp2003as:*
7: genesegp2003bs:*
8: genesegp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1659	100.0	299	2	AAW04928
2	1659	100.0	299	2	AAW63624
3	1178	71.0	305	2	AAW4854
4	1178	71.0	305	2	AAW41929
5	1159	69.9	305	2	AAW15271
6	1159	69.9	305	2	AAW25428
7	1159	69.9	305	2	AAW28295
8	1159	69.9	305	2	AAW25525
9	1159	69.9	305	2	AAW25464
10	1159	69.9	305	2	AAW27963
11	1159	69.9	305	2	AAW42063
12	1159	69.9	305	2	AAW37150
13	1159	69.9	305	2	AAW67388
14	1159	69.9	305	2	AAW01502
15	1159	69.9	305	2	AAW88471
16	1159	69.9	305	2	AAW44266
17	1159	69.9	305	2	AAW46616
18	1159	69.9	305	2	AAW03660
19	1159	69.9	305	2	AAW00014
20	1146	69.1	305	2	AAW78352
21	1146	69.0	305	2	AAW78356
22	1143	68.9	305	2	AAW78359
23	1142.5	68.9	306	2	AAW44269
24	1142	68.8	305	2	AAW28300
25	1142	68.8	305	2	AAW78355

26	1142	68.8	305	2	AAW78358	AAW78358 Humicola
27	1138	68.6	305	2	AAW78353	AAW78353 Humicola
28	1137	68.5	305	2	AAW78360	AAW78360 Humicola
29	1137	68.5	305	2	AAW78354	AAW78354 Humicola
30	1137	68.5	305	2	AAW78357	AAW78357 Humicola
31	1134	68.4	305	2	AAW28818	AAW28818 H. insol
32	1127	67.9	286	2	AAW57420	AAW57420 Humicola
33	1122	67.6	200	2	AAW53968	AAW53968 Thielavia
34	1120	67.5	284	2	AAW04137	AAW04137 Mutant 43
35	1119	67.5	284	2	AAW04129	AAW04129 Mutant 43
36	1117	67.3	200	2	AAW53979	AAW53979 Thielavia
37	1114	67.1	284	2	AAW04141	AAW04141 Mutant 43
38	1113	67.1	284	2	AAW04140	AAW04140 Mutant 43
39	1112	67.0	284	2	AAW04130	AAW04130 Mutant 43
40	1112	67.0	284	2	AAW04128	AAW04128 Humicola
41	1110	66.9	305	2	AAW78361	AAW78361 Humicola
42	1108	66.8	284	2	AAW04131	AAW04131 Mutant 43
43	1108	66.8	284	2	AAW04135	AAW04135 Mutant 43
44	1107.5	66.8	306	2	AAW44270	AAW44270 Hybrid DN
45	1107	66.7	284	2	AAW04132	AAW04132 Mutant 43

ALIGNMENTS

RESULT 1

ID AAW04928 standard; protein; 299 AA.

AC AAW04928;

DT 20-MAY-1997 (first entry)

DE Cellulytic enzyme #4 of the invention.

XX Cellulytic enzyme; endoglucanase; hydrolysis; cellulose; microorganism;

XX plant cellulase; catalytic region; textile; backstaining; bio-polishing;

XX stone-washing; cellulosic fabric; colour clarification; debarkation;

XX cell wall degradation; paper pulp; debarking; fibre modification;

XX enzymatic de-inking; drainage improvement.

XX Thielavia terrestris.

XX M09629397-AL.

XX 26-SEP-1996.

XX 18-MAR-1996; 96WO-DK00105.

XX 17-MAR-1995; 95DK-00000272.

XX 08-AUG-1995; 95DK-00000885.

XX 08-AUG-1995; 95DK-00000886.

XX 08-AUG-1995; 95DK-00000887.

XX 08-AUG-1995; 95DK-00000888.

XX 12-FEB-1996; 96DK-00000137.

XX (NOVO) NOVO-NORDISK AS.

XX Schuelein M, Andersen LN, Lasser SF, Kaupinen MS, Lange L;

XX Nielsen RI, Ihara M, Takagi S;

XX WPI: 1996-443173/44.

XX N-PSDB; AAT39050.

XX New endo-glucanase enzyme preparations - contg. conserved catalytic

XX regions, useful for treating fabrics, textiles, plant material or paper

XX pulp.

Claim 72; Page 161-163; 316pp; English.

AAW04925-W04932 represent the enzymes of the invention. These enzymes

possess cellulytic (particularly endoglucanase) activity. Cellulytic

enzymes are involved in the the hydrolysis of cellulose, and are

CC synthesised by a large number of microorganisms and plants. The enzymes
CC of the invention containing the conserved catalytic regions (such as
CC AAM04913) exhibit improved performance, e.g. 50 times higher performance,
CC compared to multiple domain enzymes. The enzymes can be used for the
CC treatment of fabrics or textiles, preferably for preventing backstaining,
CC for bio-polishing or for stone-washing cellulosic fabric. They can also
CC be used to provide colour clarification for laundry. The enzymes can also
CC be used for the degradation or modification of plant material, such as
CC cell walls. They can also be used in the treatment of paper pulp
CC preferably for debarking, defibration, fibre modification, enzymatic de-
CC laking or drainage improvement

XX Sequence 299 AA;

Query Match 100.0%; Score 1659; DB 2; Length 299;

Best Local Similarity 100.0%; Pred. No. 8,6e-119;

Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSTPVLRTTLAALPLVSAASGSGSTRYMDCKPCSCAMPKAAVSQPVYACDANFOR 60
DB 1 MRSTPVLRTTLAALPLVSAASGSGSTRYMDCKPCSCAMPKAAVSQPVYACDANFOR 60

QY 61 LSDPNVSGCMGSGSAVSCADDTPMAYNDNLAYGPAATSIAGSESSWCACALYLTSSGP 120
DB 61 LSDPNVSGCMGSGSAVSCADDTPMAYNDNLAYGPAATSIAGSESSWCACALYLTSSGP 120

QY 121 VAGKTWVVGSTSTGDLGSGNOFDIAMPGGVGIFNGCSSQFGLPGAQYGGISSRDQDS 180
DB 121 VAGKTWVVGSTSTGDLGSGNOFDIAMPGGVGIFNGCSSQFGLPGAQYGGISSRDQDS 180

QY 181 FPAPLKPQCGMRPFMFONADNPTFTFOOVQCPAEIVASGCKRNDSSFVPTPPSGNG 240
DB 181 FPAPLKPQCGMRPFMFONADNPTFTFOOVQCPAEIVASGCKRNDSSFVPTPPSGNG 240

QY 241 GTGPTSTAPSGSGTSPGSGSGCTSQKAAQCGGIFSGCTTVCSTTQCKLNDYISQCL 299
DB 241 GTGPTSTAPSGSGTSPGSGSGCTSQKAAQCGGIFSGCTTVCSTTQCKLNDYISQCL 299

RESULT 2
AAM63624
ID AAM63624 standard; protein; 299 AA.

XX AAM63624;
AC
DT 21-SEP-1998 (first entry)

XX Monocomponent endoglucanase enzyme.

XX Monocomponent endoglucanase; cellulolytic enzyme; garment; abrasion;
KW abraded looking jeans; fungus; Thielavia terrestris; stone-washed.

XX Thielavia terrestris.

XX EP843041-AL.

XX 20-MAY-1998.

XX 30-MAY-1997; 37EP-00610021.

XX 13-NOV-1996; 96DK-00001276.

XX (NOVO) NOVO-NORDISK AS.

XX Lund H, Kalum Jr.

XX WPI; 1998-263308/24.

XX N-PSDB; AAV39396.

XX Pair of jeans - with abraded look.
XX
PS Claim 16; Page 10-12; 16pp; English.
XX

CC This represents a monocomponent endoglucanase. This is a cellulolytic
CC enzyme belonging to the fungal species Thielavia terrestris. This can be
CC used in the process of invention of providing a pair of jeans made from
CC dyed twill fabric that has localised variations in the colour density of
CC the fabric. The process involves providing the jeans with a stone-washed
CC or abraded look, where the variation corresponds to a delta remission
CC value (Delta R) higher than 11, and the reflection of a first area of the
CC jeans fabric is less than 12 percent. The process comprises selecting the
CC desired textile cutting pattern for jeans garment, positioning the
CC pattern onto newly manufactured dyed twill fabric, cutting the jeans
CC garment parts, sewing the pair of jeans, optionally subjecting the pair
CC of jeans to a desizing treatment, and subjecting the pair of jeans to an
CC abrasion treatment with an efficient amount of a cellulolytic enzyme in
CC an aqueous medium essentially free of bleaching chemicals. The abraded
CC look near stitching is different from the abraded look far from stitching

XX Sequence 299 AA;

Query Match 100.0%; Score 1659; DB 2; Length 299;

Best Local Similarity 100.0%; Pred. No. 8,6e-119;

Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSTPVLRTTLAALPLVSAASGSGSTRYMDCKPCSCAMPKAAVSQPVYACDANFOR 60
DB 1 MRSTPVLRTTLAALPLVSAASGSGSTRYMDCKPCSCAMPKAAVSQPVYACDANFOR 60

QY 61 LSDPNVSGCMGSGSAVSCADDTPMAYNDNLAYGPAATSIAGSESSWCACALYLTSSGP 120
DB 61 LSDPNVSGCMGSGSAVSCADDTPMAYNDNLAYGPAATSIAGSESSWCACALYLTSSGP 120

QY 121 VAGKTWVVGSTSTGDLGSGNOFDIAMPGGVGIFNGCSSQFGLPGAQYGGISSRDQDS 180
DB 121 VAGKTWVVGSTSTGDLGSGNOFDIAMPGGVGIFNGCSSQFGLPGAQYGGISSRDQDS 180

QY 181 FPAPLKPQCGMRPFMFONADNPTFTFOOVQCPAEIVASGCKRNDSSFVPTPPSGNG 240
DB 181 FPAPLKPQCGMRPFMFONADNPTFTFOOVQCPAEIVASGCKRNDSSFVPTPPSGNG 240

QY 241 GTGPTSTAPSGSGTSPGSGSGCTSQKAAQCGGIFSGCTTVCSTTQCKLNDYISQCL 299
DB 241 GTGPTSTAPSGSGTSPGSGSGCTSQKAAQCGGIFSGCTTVCSTTQCKLNDYISQCL 299

RESULT 3
AAM44854
ID AAM44854 standard; protein; 305 AA.

XX AAM44854;

XX 31-JUL-1998 (first entry)

XX Humicola insolens cellulase NCB4 protein.

XX Humicola insolens; NCB1; NCB2; NCB4; cellulase; expression vector;

XX promoter; signal sequence; terminator; amylase; lipase; protease;

XX Humicola insolens.

XX Key Location/Qualifiers

XX Peptide 1..21

XX Protein 22..305

XX W09803667-AL.

XX 29-JAN-1998.

XX 24-JUL-1997; 97WO-JP002560.

XX 24-JUL-1996; 96UP-00195070.

2A (MEIJ) MEIJI SEIKA KAISHA LTD.
 XX
 P1 Moriya T, Murashima K, Aoyagi K, Sumida N, Watanabe M, Hamaya T;
 P1 Koga J, Kono T, Murakami T;
 DR N-PSDB; AAV19378.
 WP: 1998-120786/11.
 PT Mass production of proteins and peptides in Humicola species - using
 PT expression vector containing the promoter, signal sequence and/or
 PT terminator from the Humicola insolens NCE1 or NCE2 gene.
 XX
 PS Example A6; Page 46-48; 63pp; Japanese.
 XX
 CC The present sequence represents the Humicola insolens cellulase NCE4
 CC protein from the present invention. The present invention describes a
 CC method for the mass production of proteins and peptides in Humicola
 CC species, especially in Humicola insolens, using an expression vector
 CC which comprises the promoter, signal sequence and/or terminator
 CC regulatory sequences from the NCE1 or NCE2 gene of H. insolens. These are
 CC available in the plasmids pM3-1 (Escherichia coli JM109/pM3-1, FERM BP-
 CC 5971) (for NCE1) and pM14-1 (E. coli JM109/pM14-1, FERM BP-5972) (for
 CC NCE2). The vector also contains a marker gene such as an antibiotic
 CC resistance gene (e.g. the tetracycline resistance gene from Streptomyces
 CC rifaciens). Proteins which can be expressed using this system include
 CC cellulase, amylase, lipase, protease, phytase and other enzymes. Specific
 CC expression vectors of the invention are pMDO1 (for Humicola NCE3
 CC cellulase gene), pEGD01 (for Humicola NCE4 cellulase gene) and pMDO2
 CC (for Humicola NCE4 cellulase gene). The expression system allows the
 CC efficient production of proteins and peptides in a Humicola host. Using
 CC the expression system high amounts of protein (>4.5 g/l) can be obtained
 XX
 SQ Sequence 305 AA;
 Query Match 71.0%; Score 1178; DB 2; Length 305;
 Best Local Similarity 68.1%; Pred. No. 5.1e-82;
 Matches 209; Conservative 44; Mismatches 44; Indels 10; Gaps 4;
 QY 1 MRSTPVLRTTAAALPLIVASASGSGSTRYNDCKSPCAMPGKAAVOPVACDANFOR 60
 DB 1 MRSSPLRSANVALPVALAA--DGKSTRYNDCKSPCAMPGKAAVOPVACDANFOR 58
 QY 120 PVAGKTMVOSTTGGDLGSGNPFDIAMPGGVGIFFNGCSQFGGLPGAQYGGISRPDQC 179
 DB 119 PVAGKTMVOSTTGGDLGSGNPFDIAMPGGVGIFFNGCSQFGGLPGAQYGGISRPDQC 178
 QY 180 SFPAFLKPGCOMFDMFQANLPTFTFOYQVCPALIVASGCKRNDSSFP-VFTPPSG 238
 DB 179 RFPDLKPGCOMFDMFQANLPTFTFOYQVCPALIVASGCKRNDSSFP-VFTPPSG 238
 QY 239 NGGTGTPTSTAPGSGTS-----PGGSGCTGKMAQCGGIGSGCTTCVSGTCCOKLN 292
 DB 239 SGPVGPPTSTSTSTSTSSPPVQPTPSGCTARMAQCGGIGSGCTTCVSGTCCOKLN 298
 QY 293 DYISQCL 299
 DB 299 DMYHQCL 305
 RESULT 4
 AAW41929
 ID AAW41929 standard; protein; 305 AA.
 XX
 AC AAW41923;
 XX
 DE 26-JUN-1998 (first entry)
 XX Humicola insolens endoglucanase cellulase NCE4.
 XX

KW Endoglucanase; cellulase; NCE4; fluff removal; weight reduction;
 KW denier reduction; bleaching; denim dyed fibre.
 OS Humicola insolens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..22
 FT /label= sig_peptide
 FT Peptide 23..305
 FT /label= mat_peptide
 XX
 PN MO9803640-A1.
 XX
 PD 29-JUN-1998.
 XX
 PF 24-JUL-1997; 97MO-JD002561.
 XX
 PR 24-JUL-1996; 96JP-00194974.
 XX
 PA (MEIJ) MEIJI SEIKA KAISHA LTD.
 XX
 P1 Murashima K, Moriya T, Hamaya T, Koga J, Sumida N, Aoyagi K,
 P1 Murakami T, Kono T;
 DR N-PSDB; AAV13840.
 WP: 1998-120765/11.
 PT Endoglucanase cellulase NCE4 from Humicola insolens - for treatment of
 PT cellulose fibres for fluff elimination and weight reduction.
 XX
 PS Claim 1; Page 33-36; 43pp; Japanese.
 CC The present sequence is the Humicola insolens endoglucanase cellulase
 CC NCE4, which can be used to treat cellulose fibres for fluff removal,
 CC weight (denier) reduction and bleaching, especially of denim dyed fibres
 XX
 SQ Sequence 305 AA;
 Query Match 71.0%; Score 1178; DB 2; Length 305;
 Best Local Similarity 68.1%; Pred. No. 5.1e-82;
 Matches 209; Conservative 44; Mismatches 44; Indels 10; Gaps 4;
 QY 1 MRSTPVLRTTAAALPLIVASASGSGSTRYNDCKSPCAMPGKAAVOPVACDANFOR 60
 DB 1 MRSSPLRSANVALPVALAA--DGKSTRYNDCKSPCAMPGKAAVOPVACDANFOR 58
 QY 120 PVAGKTMVOSTTGGDLGSGNPFDIAMPGGVGIFFNGCSQFGGLPGAQYGGISRPDQC 179
 DB 119 PVAGKTMVOSTTGGDLGSGNPFDIAMPGGVGIFFNGCSQFGGLPGAQYGGISRPDQC 178
 QY 180 SFPAFLKPGCOMFDMFQANLPTFTFOYQVCPALIVASGCKRNDSSFP-VFTPPSG 238
 DB 179 RFPDLKPGCOMFDMFQANLPTFTFOYQVCPALIVASGCKRNDSSFP-VFTPPSG 238
 QY 239 NGGTGTPTSTAPGSGTS-----PGGSGCTGKMAQCGGIGSGCTTCVSGTCCOKLN 292
 DB 239 SGPVGPPTSTSTSTSTSSPPVQPTPSGCTARMAQCGGIGSGCTTCVSGTCCOKLN 298
 QY 293 DYISQCL 299
 DB 299 DMYHQCL 305
 RESULT 5
 AAR15271
 ID AAR15271 standard; protein; 305 AA.
 XX
 AC AAR15271;
 XX

QY 293 DYSOCL 299
Db 299 DWHQCL 305

RESULT 7
ID AAR28295 standard; protein; 305 AA.
AAR28295;

24-OCT-2003 (revised)
25-MAR-2003 (revised)
02-APR-1993 (first entry)

Sequence of ~ 43 kD endoglucanase.

Detergent additive; protease; enzyme; endoglucanase.

Humicola insolens; DSM 1800.

Key Location/Qualifiers

Peptide 1..21
label= signal

Protein 22..305

MO9218599-AI.

29-OCT-1992.

10-APR-1992: 92MO-DK000116.

22-APR-1991: 91DK-00000737.

(NOVO) NOVO-NORDISK AS.

Wagner P, Tsuchiya R;

WPI; 1992-382092/46.

N-PSDB; AAQ20067.

Detergent additive concy. cellulase and specific protease - which does

not degrade the cellulase during storage and clarifies the colour of dyed

cellulosic materials.

Disclosure; Page 8-9; 15pp; English.

The ~43 kD endoglucanase derived from Humicola insolens, DSM 1800 is

described in detail in co-pending Danish patent application No. DK

1159/90. As a detergent additive, it has a higher degree of specificity

than Bacillus licheniformis protease. The term "higher degree of

specificity" is defined as a protease which conditions degrades human

insulin to fewer components. (Updated on 25-MAR-2003 to correct PN

field.) (Updated on 24-OCT-2003 to standardise OS field)

Sequence 305 AA;

Query Match 69.9%; Score 1159; DB 2; Length 305;

Best Local Similarity 67.1%; Pred. No. 1.4e-80;

Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 180 SFPAELKPGQWRPFMDPQADNPTFTFOQVOCFAIVASGCKRNDSSFP-VFTPPSGG 238
Db 179 RFPALKPGQWRPFMDPQADNPTFTFOQVOCFAIVASGCKRNDSSFP-VFTPPSGG 238

QY 239 NGTGTPTSTAPSGQTS-----PGGSGCTSQKMAOCGIGFSGCTTCVSGTTCQXN 292
Db 239 SSPVQPTSTSTSTSTSTSPVQPTTPSGCTAERMAOCGNGMSGCTTCVSGTTCQXN 298

QY 293 DYSOCL 299
Db 299 DWHQCL 305

RESULT 8
ID AAR25525 standard; protein; 305 AA.
AAR25525;

24-OCT-2003 (revised)
25-MAR-2003 (revised)
11-JAN-1993 (first entry)

Humicola insolens DSM 1800 endoglucanase.

Detergent; washing powder; cellulase; softening clay; Cl4CWC-method;

flocculation; radio-labelled carboxymethylcellulase.

Humicola insolens; DSM 1800.

Key Location/Qualifiers

Peptide 1..21
label= signal

Protein 22..305
label= endoglucanase

EP495258-AI.

22-JUL-1992.

06-NOV-1991: 91EP-00202880.

16-JAN-1991: 91EP-00870006.

(PROC) PROCTER & GAMBLE CO.

Convents AC, Busch A, Baack AC;

WPI; 1992-243163/30.

N-PSDB; AAQ26405.

Compact, granular detergent compans. - contain high activity cellulase

and softening clay to provide synergistic effect in softening

performance.

Claim 5; Page 29-30; 38pp; English.

The endoglucanase isolated from H. insolens DSM 1800 has Mr 43,000 and

high cellulase activity (i.e. removes a minimum of 10% of radiolabelled

carboxymethylcellulose (Cl4CWC)). The enzyme can be used in detergent

compositions along with a surface active agent, a builder system and a

softening clay. See also AAQ26407. (Updated on 25-MAR-2003 to correct PN

field.) (Updated on 24-OCT-2003 to standardise OS field)

Sequence 305 AA;

Query Match 69.9%; Score 1159; DB 2; Length 305;

Best Local Similarity 67.1%; Pred. No. 1.4e-80;

Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSTPURTLLAALPLVSAASGSGSTRYWDCKRSCAMPKAAVSPVACDANFOR 60
Db 119 PVAGKMTVYSTSTGSLGSLNODFIAMPGGSGVGRNCCSSGSGGIPKQVYGISSSDCCD 179
119 PVAGKMTVYSTSTGSLGSLNODFIAMPGGSGVGRNCCSSGSGGIPKQVYGISSSDCCD 179

Db 1 MRSSEPLPSAVNALPVALAA--DGRSTRYMDCKPSGMAKAPVNPVPSGNANFOR 58
 QY 61 LSDPFWVQSGCN-GGSAYSCADQTPMAVNDNLAYGPAATSIAGSSBSWCCACALYFTTSG 119
 Db 59 ITDFPAKSGCBPGVAYSCADQTPMAVNDNLAYGPAATSIAGSSBSWCCACALYFTTSG 118
 QY 120 PVAGKTMVQVSTSTGDLGSDNFDLAMPGGVGTGNGSSQSGLPQAGVGGISRRQCD 179
 Db 119 PVAGKTMVQVSTSTGDLGSDNFDLAMPGGVGTGNGSSQSGLPQAGVGGISRRQCD 178
 QY 180 SFPAFLKPGQCMRPFDMFONADNPTFTFOVOCFAEIVARSCKRNDSSFP-VFTPPSG 238
 Db 179 RFPDALKPGCYRMRDMFNADNPSFSFQVOCFAEIVARTGCRNDGNFPAVQIPSSST 238
 QY 239 NGGTGPTSTAPGSQTS-----PGSGSGCTSQKMAQCGIGPSGCTTCVSGTTCOKLN 292
 Db 239 SSPVNGPTSTSTSTSTSSPVPQPTTPSGCTAERMAQCGGNGSGCTTCVAGSTCTKIN 298
 QY 293 DYISQCL 299
 Db 299 DMHQCL 305

RESULT 9

AAR25464
 ID AAR25464 standard; protein; 305 AA.

XX AC AAR25464;
 XX DT 25-MAR-2003 (revised)
 XX DT 07-JAN-1993 (first entry)
 XX DE Endoglucanase #1.
 XX KM CMC-endoase; 43 kD cellulase; monoclonal antibody.
 XX OS Humicola insolens.
 XX FH Key Location/Qualifiers
 XX FT Peptide 1..21
 XX FT Protein /label=Signal_peptide
 XX FT Protein 22..305
 XX FT Protein /label=Mature_protein
 XX PN EP495257-A1.
 XX PD 22-JUL-1992.
 XX PF 06-NOV-1991; 91BP-00202879.
 XX PR 16-JAN-1991; 91EP-00870006.
 XX PA (PROC) PROCTER & GAMBLE CO.
 XX PI Baeck AC, Busch A, Ceulemans FAA;
 XX DR MPI; 1992-243162/30.
 XX DR N-PSDB; AAQ26380.
 XX PT Compact, granular detergent compns. - contain high activity cellulase
 XX PT and softening clay to provide synergistic effect in softening
 XX PT performance.
 XX PS Disclosure; Page 20-21; 29pp; English.
 XX CC The sequences given in AAR25464 and AAR25466 are endoglucanases which are
 XX CC immunoreactive with a monoclonal antibody raised against a partially
 XX CC purified 43 kD cellulase derived from Humicola insolens. These
 XX CC endoglucanases exhibit a CMC-endoase activity of at least about 50, pref.
 XX CC at least about 60, imparticular at least about 90 CMC- endase units per
 XX CC mg of total protein. These endoglucanases have molecular weight of
 XX CC approx. 43 kD. (Updated on 25-MAR-2003 to correct PW field.) (Updated on
 XX CC 25-MAR-2003 to correct DR field.)

XX Sequence 305 AA.

Query Match 69.9%; Score 1159; DB 2; Length 305;
 Best Local Similarity 67.1%; Pred. No. 1,4e-80;
 Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSPTVRLTTLAAALPVALASAGSGSFRYMDCKPSGMAKAPVNPVPSGNANFOR 60
 Db 1 MRSSEPLPSAVNALPVALAA--DGRSTRYMDCKPSGMAKAPVNPVPSGNANFOR 58
 QY 61 LSDPFWVQSGCN-GGSAYSCADQTPMAVNDNLAYGPAATSIAGSSBSWCCACALYFTTSG 119
 Db 59 ITDFPAKSGCBPGVAYSCADQTPMAVNDNLAYGPAATSIAGSSBSWCCACALYFTTSG 118
 QY 120 PVAGKTMVQVSTSTGDLGSDNFDLAMPGGVGTGNGSSQSGLPQAGVGGISRRQCD 179
 Db 119 PVAGKTMVQVSTSTGDLGSDNFDLAMPGGVGTGNGSSQSGLPQAGVGGISRRQCD 178
 QY 180 SFPAFLKPGQCMRPFDMFONADNPTFTFOVOCFAEIVARSCKRNDSSFP-VFTPPSG 238
 Db 179 RFPDALKPGCYRMRDMFNADNPSFSFQVOCFAEIVARTGCRNDGNFPAVQIPSSST 238
 QY 239 NGGTGPTSTAPGSQTS-----PGSGSGCTSQKMAQCGIGPSGCTTCVSGTTCOKLN 292
 Db 239 SSPVNGPTSTSTSTSTSSPVPQPTTPSGCTAERMAQCGGNGSGCTTCVAGSTCTKIN 298
 QY 293 DYISQCL 299
 Db 299 DMHQCL 305

RESULT 10

AAR27968
 ID AAR27968 standard; protein; 305 AA.

XX AC AAR27968;
 XX DT 24-OCT-2003 (revised)
 XX DT 25-MAR-2003 (revised)
 XX DT 09-MAR-1993 (first entry)
 XX DE Endoglucanase enzyme.
 XX KM Alkaline cellulase; laundry detergent compositions;
 XX KM fabric colour maintenance.
 XX OS Humicola insolens; DSM 1800.
 XX FH Key Location/Qualifiers
 XX FT Peptide 1..21
 XX FT Protein /note="signal peptide"
 XX FT Peptide 22..305
 XX FT Protein /note="mature peptide"
 XX PN 3P508358-A1.
 XX PD 14-OCT-1992.
 XX PF 07-APR-1992; 92BP-00105956.
 XX PR 12-APR-1991; 91EP-00870062.
 XX PA (PROC) PROCTER & GAMBLE CO.
 XX PI Busch A, Maccorquodale F;
 XX DR MPI; 1992-341667/42.
 XX DR N-PSDB; AAQ23934.
 XX PT Laundry detergent compns. - contg. alkali cellulase and FVP for fabric
 XX PT colour maintenance upon laundering.

PS Disclosure; Page 16; 23pp; English.

XX The sequence is that of an endoglucanase (alkaline cellulase) from

CC Humicola insolens which is used with a polyvinylpyrrolidone (PVP) in a

CC laundry detergent compsn. The action of the PVP and the endoglucanase in

CC fabric colour maintenance upon laundering is superior to the sum of the

CC individual actions of both these ingredients. See also AAR27699. (Updated

CC on 25-MAR-2003 to correct PN field.) (Updated on 24-OCT-2003 to

CC standardise OS field)

XX

XX Sequence 305 AA;

SQ

Query Match 69.9%; Score 1159; DB 2; Length 305;

Best Local Similarity 67.1%; Pred. No. 1.4e-80;

Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 KRSTPVLRFTTLAALPLVASAASGSGSTRYDCCPCSCAMPKAAVSGPVYACDANFOR 60

DB : KRSSPLPSAVVALPVALALAA--DKRSTRYDCCPCSCMAKKAAPVNPQVPSGNANFOR 58

QY 61 LSDPFWVSGCN-GGSAVSCADQTPMAVNDNLAYGFATSIAGSSSSWCCACVALTFTSG 119

DB 59 ITDFPAKSGCEPGVAVSCADQTPMAVNDNLAYGFATSIAGSSSSWCCACVALTFTSG 118

QY 120 PVAGKTMVVGSTSTGDLGSGNFDIAMEGGVGIENGSSQFGILPQAYGGISSRDQD 179

DB 119 PVAGKTMVVGSTSTGDLGSGNFDIAMEGGVGIENGSSQFGILPQAYGGISSRDQD 178

QY 180 SFPAFLKPGCGQMRDFQANADPTTFQVOCPEAIVARSCKRNDSSFP-VFTPPSGG 238

DB 179 RFPDAKPGCYRDFQKADNPSPFQVQCPAE--NARTGCRNDGNFPVAVQIPSSST 238

QY 239 NGGTGTPTSTAPSGQTS-----PGSGSGCTSGKMAQCGGIGFSGGCTTGVSGTTCQKLN 292

DB 239 SSPVNPPTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGGNGSGCTTCVAGSTCTKLN 298

QY 233 DYSQCL 299

DB 299 DMYHQL 305

RESULT 11

AAR42063

ID AAR42063 standard; protein; 305 AA.

XX AAR42063;

AC

DT 28-APR-1994 (first entry)

XX

DE Endoglucanase enzyme.

XX

XX Detergent; fabric; surfactant; softening clay; cellulase;

KW Humicola insolens; Fusarium oxysporum; endoglucanase.

XX

OS Humicola insolens.

XX

PH Key Location/Qualifiers

FT Peptide 1..21

FT /label= sig_peptide

FT Protein 22..305

FT /label= mat_protein

XX

PN AU9211048-A.

XX

PD 02-SEP-1993.

XX

PF 18-FEB-1992; 32AU-00011048.

XX

PR 18-FEB-1992; 32AU-00011048.

XX

PA (PROC) PROCTER & GAMBLE CO.

XX

PI Corvents AC, Busch A, Baech AC;

XX WPI; 1993-328419/42.

DR N-PSDB; AA049941.

XX

PT Detergent compsn., esp. for fabrics - contain surfactant, softening clay

PT and high activity cellulase, partic. from Humicola insolens.

XX

PS Claim 5; Page 54-55; 71pp; English.

XX

CC A new detergent comprises a high activity cellulase in combination with a

CC softening clay. The cellulase may be an endoglucanase enzyme derived from

CC Humicola insolens (AA049941) or Fusarium oxysporum (AA049942). The

CC combination provides a higher than additive softening performance and

CC excellent colour rejuvenation and whiteness maintenance for fabrics

XX

XX Sequence 305 AA;

SQ

Query Match 69.9%; Score 1159; DB 2; Length 305;

Best Local Similarity 67.1%; Pred. No. 1.4e-80;

Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 KRSTPVLRFTTLAALPLVASAASGSGSTRYDCCPCSCAMPKAAVSGPVYACDANFOR 60

DB 1 KRSSPLPSAVVALPVALALAA--DKRSTRYDCCPCSCMAKKAAPVNPQVPSGNANFOR 58

QY 61 LSDPFWVSGCN-GGSAVSCADQTPMAVNDNLAYGFATSIAGSSSSWCCACVALTFTSG 119

DB 59 ITDFPAKSGCEPGVAVSCADQTPMAVNDNLAYGFATSIAGSSSSWCCACVALTFTSG 118

QY 120 PVAGKTMVVGSTSTGDLGSGNFDIAMEGGVGIENGSSQFGILPQAYGGISSRDQD 179

DB 119 PVAGKTMVVGSTSTGDLGSGNFDIAMEGGVGIENGSSQFGILPQAYGGISSRDQD 178

QY 180 SFPAFLKPGCGQMRDFQANADPTTFQVOCPEAIVARSCKRNDSSFP-VFTPPSGG 238

DB 179 RFPDAKPGCYRDFQKADNPSPFQVQCPAE--NARTGCRNDGNFPVAVQIPSSST 238

QY 239 NGGTGTPTSTAPSGQTS-----PGSGSGCTSGKMAQCGGIGFSGGCTTGVSGTTCQKLN 292

DB 239 SSPVNPPTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGGNGSGCTTCVAGSTCTKLN 298

QY 233 DYSQCL 299

DB 299 DMYHQL 305

RESULT 12

AAR37150

ID AAR37150 standard; protein; 305 AA.

XX AAR37150;

AC

DT 25-MAR-2003 (revised)

DT 25-AUG-1993 (first entry)

XX

DE Dye transfer inhibiting compsn. cellulase.

XX

XX Detergent; homogeneous endoglucanase component; 43kD cellulase.

KW Humicola insolens.

XX

OS Humicola insolens.

XX

PH Key Location/Qualifiers

FT Peptide 1..21

FT /note= "signal peptide"

FT Peptide 22..305

FT /note= "mature peptide"

XX

PN EP540784-A1.

XX

PD 12-MAY-1993.

XX

PF 06-NOV-1991; 91EP-00202882.

PR 36-NOV-1991; 91EP-00202882.
 XX (PROC) PROCTER & GAMBLE CO.
 XX
 XX Mccorquodale F, Busch A;
 XX
 XX WPI: 1993-153688/19.
 DR N-PSDB: AA041732.
 XX
 XX Compn. for inhibiting dye transfer during fabric washing - contains
 PT peroxidase, hydrogen peroxide, substrate and cellulase, esp. endo-
 PT glucanase from Humicola insolens.
 XX
 XX Disclosure; Page 18-19; 28pp; English.
 XX
 XX The sequence is that of a cellulase which is characterised in that the
 CC cellulase provides at least 10% removal of immobilised radioactive
 CC labelled carboxymethyl cellulose according to the C14CMC method at 25 x
 CC 10(-6) by weight of the cellulase protein in the test solution. It can
 CC be used as part of a compn. for inhibiting dye transfer. (Updated on 25-
 CC MAR-2003 to correct PN field.)
 XX
 XX SQ Sequence 305 AA;
 Query Match 69.9%; Score 1159; DB 2; Length 305;
 Best Local Similarity 67.1%; Pred. No. 1.4e-80;
 Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;
 QY 1 MRSTPVLRITTLAALPLVSAASGSGSTRYWDCKRSCAMPKGAASQPVYACDANFOR 60
 DB 1 MRSSPLPSAVVAALPVYALAA--DGRSTRYWDCKRSCGMAKAPVNFSCNANFOR 58
 QY 61 LSPDNVOSGCN-GGSAYSCADQTPMAVNDNLAYGFATSIAGSSSSWCCATLFTTSG 119
 DB 59 ITDFDAKSGCGEPGVAVSCADQTPMAVNDNLAYGFATSIAGSSSSWCCATLFTTSG 118
 QY 120 PVAGKMWVOSTSTGGDLGSGNDFIAMPGGGVGIFNGCSSQPGGLPCAQYGGISRPDCC 179
 DB 119 PVAGKMWVOSTSTGGDLGSGNDFIAMPGGGVGIFNGCSSQPGGLPCAQYGGISRPDCC 178
 QY 180 SEPAPLKPCCWRPFQNDPNTFTFPQVQCPAEIVASGCKRNDSSPP-VFTPPSGG 238
 DB 179 RFPDALKPGCYWRPFQNDPNTFTFPQVQCPAEIVASGCKRNDSSPP-VFTPPSGG 238
 QY 239 NGGTGTPTSTAPSGGTS-----PGSGSGCTGKMAQCGGIGFSGCTCVSGTGOXKN 292
 DB 239 SSPVQPTSTSTSTSTSTSPVQPTTPSGCTAERMAQCGGNGSGCTTCVAGSTCTKIN 298
 QY 293 DYVSQCL 299
 DB 299 DMTHQCL 305
 QY
 DB
 RESULT 13
 AA067388
 ID AA067388 standard; protein; 305 AA.
 XX
 XX AA067388;
 XX
 XX 25-APR-2000 (first entry)
 XX
 XX H. insolens endoglucanase enzyme protein sequence.
 DE
 XX
 XX Endoglucanase; cellulase enzyme; detergent composition; anion surfactant;
 KW degradation protection; liquid detergent; long term stability.
 XX
 XX Humicola insolens.
 OS
 XX
 XX EP633311-A1.
 PV
 XX
 XX 11-JUN-1995.
 PD
 XX
 XX 28-JUN-1993; 93EP-00870122.
 PF

XX
 PR 28-JUN-1993; 93EP-00870122.
 XX
 XX (PROC) PROCTER & GAMBLE CO.
 XX
 XX Herbolts IMAU, Jansen MP;
 PI
 XX
 XX WPI: 1995-038508/06.
 DR N-PSDB: AA260178.
 XX
 XX liq. detergent compns. - contg. hydrophobic amine(s) for cellulase
 PT stabilisation.
 PT
 XX
 XX Claim 11; Page 18-19; 28pp; English.
 XX
 XX This sequence represents the Humicola insolens endoglucanase enzyme
 CC protein sequence. The enzyme is a cellulase enzyme used in the liquid
 CC detergent composition of the invention. The detergent comprises an anion
 CC surfactant, a cellulase enzyme and a cellulase stabilising amount of an
 CC amine of the formula R1R2R3N; where R1 and R2 are independently H or a C2
 CC -C9 alkyl chain, and R3 is a C2-C9 alkyl chain or cyclohexyl or
 CC cyclopentyl or cycloheptyl. The amine prevents the cellulase enzyme from
 CC becoming degraded. The composition is used as a liquid detergent, and has
 CC long term stability, it is applicable to the protection of an cellulase
 CC and can be used in the presence of proteases
 XX
 XX SQ Sequence 305 AA;
 Query Match 69.9%; Score 1159; DB 2; Length 305;
 Best Local Similarity 67.1%; Pred. No. 1.4e-80;
 Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;
 QY 1 MRSTPVLRITTLAALPLVSAASGSGSTRYWDCKRSCAMPKGAASQPVYACDANFOR 60
 DB 1 MRSSPLPSAVVAALPVYALAA--DGRSTRYWDCKRSCGMAKAPVNFSCNANFOR 58
 QY 61 LSPDNVOSGCN-GGSAYSCADQTPMAVNDNLAYGFATSIAGSSSSWCCATLFTTSG 119
 DB 59 ITDFDAKSGCGEPGVAVSCADQTPMAVNDNLAYGFATSIAGSSSSWCCATLFTTSG 118
 QY 120 PVAGKMWVOSTSTGGDLGSGNDFIAMPGGGVGIFNGCSSQPGGLPCAQYGGISRPDCC 179
 DB 119 PVAGKMWVOSTSTGGDLGSGNDFIAMPGGGVGIFNGCSSQPGGLPCAQYGGISRPDCC 178
 QY 180 SEPAPLKPCCWRPFQNDPNTFTFPQVQCPAEIVASGCKRNDSSPP-VFTPPSGG 238
 DB 179 RFPDALKPGCYWRPFQNDPNTFTFPQVQCPAEIVASGCKRNDSSPP-VFTPPSGG 238
 QY 239 NGGTGTPTSTAPSGGTS-----PGSGSGCTGKMAQCGGIGFSGCTCVSGTGOXKN 292
 DB 239 SSPVQPTSTSTSTSTSTSPVQPTTPSGCTAERMAQCGGNGSGCTTCVAGSTCTKIN 298
 QY 293 DYVSQCL 299
 DB 299 DMTHQCL 305
 QY
 DB
 RESULT 14
 AA001502
 ID AA001502 standard; protein; 305 AA.
 XX
 XX AA001502;
 XX
 XX 16-OCT-2003 (revised)
 DT 27-AUG-2003 (revised)
 DT 26-FEB-1997 (first entry)
 XX
 XX 43 kD endoglucanase, BG V.
 DE
 XX
 XX Detergent composition; cellulase; retaining-type activity;
 KW catalytic activity; cellulosiose; particulate soil removal;
 KW colour clarification; cleaning; cellulose-containing fabric;
 KW cellulohydrolase; endoglucanase.

XX Humicola insolens; DSM 1800.
 OS
 XX Key
 XX Peptide 1. .21
 FT /label= signal_peptide
 FT Protein 22. .305
 FT /label= mature_protein
 XX
 XX MO9502675-A1.
 XX
 XX 26-JAN-1995.
 XX
 XX 07-JUL-1994; 94WO-DK000280.
 XX
 XX 12-JUL-1993; 93EP-00870131.
 XX PR 11-OCT-1993; 93DK-00001135.
 XX
 XX (NOVO) NOVO-NORDISK AS.
 XX (PROC) PROCTER & GAMBLE CO.
 XX
 XX Schuelein M, Convents A, Jeffreys B, Tikhomirov DF,
 XX WPI; 1995-067325/09.
 XX
 XX Detergent compsn. contg. two cellulase components - the first removing
 XX soil particles and the second capable of colour clarification, useful in
 XX laundry compsns.
 XX
 XX Claim 25; Page 66-67; 83pp; English.
 XX
 XX Detergent compositions comprising: 1) a first cellulase component having
 XX retaining-type activity, pref. having a catalytic activity on cellulose
 XX at pH 8.5 corresponding to kcat of at least 0.01 s⁻¹, and capable of
 XX particulate soil removal; and 2) a second cellulase component having
 XX multiple domains comprising at least one non-catalytic domain attached
 XX to a catalytic domain, pref. having a catalytic activity on Red Avicel 7.5
 XX per 1 mg of cellulase protein higher than 10⁻⁴ IU and being capable of
 XX colour clarification, where at least one of the cellulase components is a
 XX single (recombinant) component, are useful for cleaning and colour
 XX clarification of cellulose-containing fabrics. The second cellulase
 XX component can be an endoglucanase which is immunoreactive with an
 XX antibody raised against a highly purified ~43 kD endoglucanase derived
 XX from Humicola insolens, DSM 1800, and is esp. the present sequence,
 XX designated EG V. (Updated on 27-AUG-2003 to correct OS field.) (Updated
 XX on 16-OCT-2003 to standardise OS field)
 XX
 XX Sequence 305 AA;
 XX
 XX Query Match 69.9%; Score 1159; DB 2; Length 305;
 XX Best Local Similarity 67.1%; Pred. No. 1.4e-80;
 XX Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;
 XX
 QY 1 MRSPTVLTITLAAALPLVSAASGSGSTRWDCKPCSCAMPAGKAASQPYACDANFOR 60
 DB 1 MRSPTLPSAVVAALPVLAALAA--DGRSTRWDCKPCSCAKKAPVNPQVFSKANFOR 58
 XX
 QY 61 LSDPNVSGCN--GSAVSCADQTPMAVNDNLAVFPAATSIAGSSSSWCCACATLTFTSG 119
 DB 59 IDTFPAKSCCEPGVAVVASCADQTPMAVNDNLAVFPAATSIAGSSSSWCCACATLTFTSG 118
 XX
 QY 120 PVAGKTMVVGSTSTGDLGSGNFDIAMPFGGSGVGIENGCSGQFGCLPGAQYGGIISRDQD 179
 DB 119 PVAGKTMVVGSTSTGDLGSGNFDIAMPFGGSGVGIENGCSGQFGCLPGAQYGGIISRDQD 178
 XX
 QY 180 SEPAPLKGCGCMRPEWQANUNFTFPQOVCPREIYVARSGGKXNDSSFP-VFTPPSGG 238
 DB 179 RFPDAKPECTWRFDWFKKNDNSFSFQVOCPELVAITGCRKRDNDNFPAYQISSST 238
 XX
 QY 239 NGGTGPTSTIAPSGSGQTS-----PQGGSGCTSCKMAQCGGIGFSGCTTCVSGTTQCKIN 292
 DB 239 SSPVNPSTSTSTSTSTSSPPVQPTTSTSGCTAFRMAQCGGSGMGSGCTTCVAGSTCKIN 298
 XX

QY 293 DYSOCL 299
 DB 299 DWTQCL 305
 XX
 XX RESULT 15
 XX AAR88471
 XX ID AAR88471 standard; protein; 305 AA.
 XX
 XX AAR88471;
 XX
 XX 16-OCT-2003 (revised)
 XX DT 03-APR-1996 (first entry)
 XX
 XX Alkaline endoglucanase Carezyme.
 XX
 XX Alkaline endoglucanase; carezyme; cellulase; host cell;
 XX KW Fusarium graminearum; heterologous gene expression.
 XX
 XX Humicola insolens; DSM 1800.
 XX
 XX Key
 XX FT Peptide 1. .21
 XX FT /label= Sig_peptide
 XX
 XX MO9600787-A1.
 XX
 XX 15-JUN-1995; 95WO-US007743.
 XX
 XX 30-JUN-1994; 94US-00269449.
 XX PR 15-MAR-1995; 95US-00404678.
 XX
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 XX
 XX Royer JC, Moyer DL, Yoder W, Shuster JR;
 XX WPI; 1996-077498/08.
 XX DR N-PSDB; AAT10182.
 XX
 XX Non-toxic, non-toxicogenic, non-pathogenic recombinant Fusarium host cell -
 XX used to produce heterologous proteins, pref. enzymes, hormones, growth
 XX factors or receptors.
 XX
 XX Claim 13; Page 22-24; 38pp; English.
 XX
 XX The Humicola insolens DSM 1800 alkaline endoglucanase Carezyme
 XX (AAR88471), a derivative of a 43 kDa cellulase, is expressed in non-
 XX toxic, non-toxicogenic, non-pathogenic, protease-deficient recombinant
 XX Fusarium graminearum ATCC 20334 carrying plasmid pDM151. This plasmid
 XX includes the Carezyme gene (AAT10182) operably linked to promoter and
 XX terminator (AAT10184-85) sequences of the Fusarium oxysporum trypsin-like
 XX protease SP387 gene. Recombinant enzyme is produced at levels of 6.0 g/l
 XX host cell; over 90% of secreted protein is Carezyme. (Updated on 16-OCT-
 XX 2003 to standardise OS field)
 XX
 XX Sequence 305 AA;
 XX
 XX Query Match 69.9%; Score 1159; DB 2; Length 305;
 XX Best Local Similarity 67.1%; Pred. No. 1.4e-80;
 XX Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;
 XX
 QY 1 MRSPTVLTITLAAALPLVSAASGSGSTRWDCKPCSCAMPAGKAASQPYACDANFOR 60
 DB 1 MRSPTLPSAVVAALPVLAALAA--DGRSTRWDCKPCSCAKKAPVNPQVFSKANFOR 58
 XX
 QY 61 LSDPNVSGCN--GSAVSCADQTPMAVNDNLAVFPAATSIAGSSSSWCCACATLTFTSG 119
 DB 59 IDTFPAKSCCEPGVAVVASCADQTPMAVNDNLAVFPAATSIAGSSSSWCCACATLTFTSG 118
 XX
 QY 120 PVAGKTMVVGSTSTGDLGSGNFDIAMPFGGSGVGIENGCSGQFGCLPGAQYGGIISRDQD 179
 DB 119 PVAGKTMVVGSTSTGDLGSGNFDIAMPFGGSGVGIENGCSGQFGCLPGAQYGGIISRDQD 178
 XX

```

Db      119 PVAGKRWVQSTSTGDLGSNHPDLNIPGGVGIFDGCTPQPGGLPGQRXGSSSRNECD 178
Qy      126 SFPALPBGQWRFPMFQONADNPTFTFOYQCPAEIVARSCGRNDSSFP-VFTPPSGG 238
      179 RFPDLPBGQWRFPMFQONADNPSFSFRQYCPAEIVARTGCRNDGNGFPAVQIPSSST 238
Qy      239 NGGCTPTSTAPSGGOTS-----PGGSGCTGOKWACCGGIGPSGCTTCVSGTTCKLN 232
Db      239 SSPVNPPTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGANGMSGCTTCVAGSTCTKIN 238
Qy      293 DYESQCL 299
      299 LMYHQL 305

```

Search completed: July 7, 2004, 08:49:38
 Job time : 61 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 08:46:10 ; Search time 22 Seconds

(without alignments)
701.644 Million cell updates/sec

Title: US-10-007-521-12

Perfect score: 1659

Sequence: 1 MRSFVRLTIAALPLVSAASGSGSTRYMDCKPSCAMPGRAAVSQPVYACDANFOR 299

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Issued Parents AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCUS.COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

2red. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1659	100.0	299	2	US-08-872-437-2
2	1659	100.0	299	3	US-08-651-136C-12
3	1659	100.0	299	4	US-09-229-911A-6
4	1178	71.0	305	3	US-09-230-222-1
5	1159	69.9	305	1	US-08-090-013-2
6	1159	69.9	305	1	US-08-081-328-2
7	1159	69.9	305	1	US-08-232-249-2
8	1159	69.9	305	2	US-08-921-426-8
9	1159	69.9	305	2	US-08-833-642A-2
10	1159	69.9	305	2	US-08-140-008A-4
11	1159	69.9	305	2	US-08-836-340-1
12	1159	69.9	305	2	US-08-389-423-2
13	1159	69.9	305	3	US-08-816-915-8
14	1159	69.9	305	3	US-09-189-060B-56
15	1159	69.9	305	4	US-09-230-665-2
16	1159	69.9	305	4	US-09-189-028-2
17	1159	69.9	305	5	PCT-US35-07743-8
18	1142.5	68.9	306	3	US-09-189-060B-66
19	1127	67.9	286	3	US-09-254-733-3
20	1125	67.8	285	4	US-09-230-225B-6
21	1112	67.0	284	1	US-08-411-777-10
22	1112	67.0	284	3	US-09-057-088-10
23	1107.5	66.8	306	3	US-09-189-060B-68
24	1076.5	64.9	308	3	US-09-189-060B-70
25	1076.5	64.9	304	3	US-09-189-060B-72
26	1045	63.0	307	3	US-09-189-060B-74
27	1039	62.6	235	3	US-09-329-350-31

28	1021	61.5	308	3	US-08-651-136C-6	Sequence 6, Appl1
29	1021	61.5	308	4	US-09-229-911A-6	Sequence 6, Appl1
30	994.5	59.9	297	3	US-08-651-136C-4	Sequence 4, Appl1
31	994.5	59.9	297	4	US-09-229-911A-4	Sequence 4, Appl1
32	968.5	58.4	234	4	US-09-230-665-6	Sequence 6, Appl1
33	962	58.0	310	3	US-08-651-136C-22	Sequence 22, Appl1
34	962	58.0	310	4	US-09-229-911A-22	Sequence 22, Appl1
35	951	57.3	295	3	US-08-651-136C-8	Sequence 8, Appl1
36	951	57.3	295	4	US-09-229-911A-8	Sequence 8, Appl1
37	939.5	56.6	376	1	US-08-090-013-4	Sequence 4, Appl1
38	939.5	56.6	376	1	US-08-081-328-4	Sequence 4, Appl1
39	939.5	56.6	376	1	US-08-232-249-4	Sequence 4, Appl1
40	939.5	56.6	376	2	US-08-833-642A-4	Sequence 4, Appl1
41	939.5	56.6	376	2	US-08-389-423-4	Sequence 4, Appl1
42	939.5	56.6	376	4	US-09-189-028-4	Sequence 4, Appl1
43	934.5	56.3	376	4	US-09-230-665-4	Sequence 4, Appl1
44	923.5	55.7	294	3	US-08-651-136C-24	Sequence 24, Appl1
45	923.5	55.7	294	4	US-09-229-911A-24	Sequence 24, Appl1

ALIGNMENTS

RESULT 1
US-08-872-437-2
Sequence 2, Application US/08872437

Patent No. 5958082

GENERAL INFORMATION:

APPLICANT: Lund, Henrik

APPLICANT: Kallum, Lisbeth

TITLE OF INVENTION: Garments With Considerable Variation In

FILE REFERENCE: 4888, 200-US

CURRENT APPLICATION NUMBER: US/08/872,437

CURRENT FILING DATE: 1997-06-10

EARLIER APPLICATION NUMBER: 1276/96

EARLIER FILING DATE: 1996-11-13

NUMBER OF SEQ ID NOS: 2

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2

LENGTH: 299

TYPE: PRT

ORGANISM: Thielavia terrestris

US-08-872-437-2

Query Match 100.0%; Score 1659; DB 2; Length 299;

Best Local Similarity 100.0%; Pred. No. 2, 5e-127; Mismatches 0; Indels 0; Gaps 0;

Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MRSFVRLTIAALPLVSAASGSGSTRYMDCKPSCAMPGRAAVSQPVYACDANFOR	60
DB	1	MRSFVRLTIAALPLVSAASGSGSTRYMDCKPSCAMPGRAAVSQPVYACDANFOR	60
QY	61	LSDFVSGNGGSAVSCADQTPMAVNDNLAYGFAATSIAGSSSSMCCATLTTSGP	120
DB	61	LSDFVSGNGGSAVSCADQTPMAVNDNLAYGFAATSIAGSSSSMCCATLTTSGP	120
QY	121	VAGKMMVQSTNGSGDLSNCFDIAMPGGVGFNCCSSQFSGI.PGAQYGISRRDQDS	180
DB	121	VAGKMMVQSTNGSGDLSNCFDIAMPGGVGFNCCSSQFSGI.PGAQYGISRRDQDS	180
QY	121	VAGKMMVQSTNGSGDLSNCFDIAMPGGVGFNCCSSQFSGI.PGAQYGISRRDQDS	180
DB	121	VAGKMMVQSTNGSGDLSNCFDIAMPGGVGFNCCSSQFSGI.PGAQYGISRRDQDS	180
QY	181	FPAPLKPCCQMRDMPONADNPTFTFOOVCPAEIYVARSQKRRDSSFFVTPPGANG	240
DB	181	FPAPLKPCCQMRDMPONADNPTFTFOOVCPAEIYVARSQKRRDSSFFVTPPGANG	240
QY	241	GTGFTSTAPSGSGTSPGSSGSGTSGVMAACGGTIGSGCTTCUSGTTCCQLNDYISQCL	299
DB	241	GTGFTSTAPSGSGTSPGSSGSGTSGVMAACGGTIGSGCTTCUSGTTCCQLNDYISQCL	299
QY	241	GTGFTSTAPSGSGTSPGSSGSGTSGVMAACGGTIGSGCTTCUSGTTCCQLNDYISQCL	299
DB	241	GTGFTSTAPSGSGTSPGSSGSGTSGVMAACGGTIGSGCTTCUSGTTCCQLNDYISQCL	299

RESULT 2
US-08-651-136C-12
Sequence 12, Application US/08651136C

Patent No. 6001639
GENERAL INFORMATION:
APPLICANT: Schulin, Martin
APPLICANT: Andersen, Lene N.
APPLICANT: Lassen, Soren F.
APPLICANT: Kauppinen, Markus S.
APPLICANT: Lange, Lene
APPLICANT: Nielsen, Ruby I.
APPLICANT: Ihara, Michiko
APPLICANT: Takagi, Shinobu
TITLE OF INVENTION: No. 6001639e1 Endoglucanases
NUMBER OF SEQUENCES: 109
CORRESPONDENCE ADDRESS:
ADDRESS: No. 6001639e1 No. 6001639e1 of No. 6001639e1 America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/651,136C
FILING DATE: 21-May-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4366,200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 299 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-651-136C-12

Query Match 100.0%; Score 1659; DB 3; Length 299;
Best Local Similarity 100.0%; Pred. No. 2.5e-127;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSPPVLTTLAALPLVASAASGSGSTRYMDCKPSCAMPKAAVSQPVYACDANFOR 60
DB 1 MRSPPVLTTLAALPLVASAASGSGSTRYMDCKPSCAMPKAAVSQPVYACDANFOR 60

QY 61 LSDPNVQSGNGSGSAYSCADQTPMAVNDNLAYGPAATSIAGSSBSMCCACVLTFTSGP 120
DB 61 LSDPNVQSGNGSGSAYSCADQTPMAVNDNLAYGPAATSIAGSSBSMCCACVLTFTSGP 120

QY 121 VAGKTMVQSTSTGDLGSGNFDIAMPGGVGI FNGCSSQFGGLPGAQYGGISRRQCD 180
DB 121 VAGKTMVQSTSTGDLGSGNFDIAMPGGVGI FNGCSSQFGGLPGAQYGGISRRQCD 180

QY 221 VAGKTMVQSTSTGDLGSGNFDIAMPGGVGI FNGCSSQFGGLPGAQYGGISRRQCD 180
DB 221 VAGKTMVQSTSTGDLGSGNFDIAMPGGVGI FNGCSSQFGGLPGAQYGGISRRQCD 180

QY 181 FPAALPGQCMRFDMFQNDNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPSGANG 240
DB 181 FPAALPGQCMRFDMFQNDNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPSGANG 240

QY 241 GTGTPSTAPSGSGTSPGGSGCTSQKMAQCGGIGFSGCTTGVSGTTQKLNYYSQL 299
DB 241 GTGTPSTAPSGSGTSPGGSGCTSQKMAQCGGIGFSGCTTGVSGTTQKLNYYSQL 299

RESULT 3
US-09-229-911A-12
Sequence 12, Application US/09229911A
Patent No. 6387690
GENERAL INFORMATION:

APPLICANT: Schulin, Martin
Andersen, Lene N.
Lassen, Soren F.
Kauppinen, Markus S.
Lange, Lene
Nielsen, Ruby I.
Ihara, Michiko
Takagi, Shinobu
TITLE OF INVENTION: No. 6387690e1 Endoglucanases
NUMBER OF SEQUENCES: 109
CORRESPONDENCE ADDRESS:
ADDRESS: No. 6387690e1 No. 6387690e1 of No. 6387690e1 America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/229,911A
FILING DATE: 13-Jan-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/651,136
FILING DATE: 21-May-1996
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4366,200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 299 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-229-911A-12

Query Match 100.0%; Score 1659; DB 4; Length 299;
Best Local Similarity 100.0%; Pred. No. 2.5e-127;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSPPVLTTLAALPLVASAASGSGSTRYMDCKPSCAMPKAAVSQPVYACDANFOR 60
DB 1 MRSPPVLTTLAALPLVASAASGSGSTRYMDCKPSCAMPKAAVSQPVYACDANFOR 60

QY 61 LSDPNVQSGNGSGSAYSCADQTPMAVNDNLAYGPAATSIAGSSBSMCCACVLTFTSGP 120
DB 61 LSDPNVQSGNGSGSAYSCADQTPMAVNDNLAYGPAATSIAGSSBSMCCACVLTFTSGP 120

QY 121 VAGKTMVQSTSTGDLGSGNFDIAMPGGVGI FNGCSSQFGGLPGAQYGGISRRQCD 180
DB 121 VAGKTMVQSTSTGDLGSGNFDIAMPGGVGI FNGCSSQFGGLPGAQYGGISRRQCD 180

QY 221 VAGKTMVQSTSTGDLGSGNFDIAMPGGVGI FNGCSSQFGGLPGAQYGGISRRQCD 180
DB 221 VAGKTMVQSTSTGDLGSGNFDIAMPGGVGI FNGCSSQFGGLPGAQYGGISRRQCD 180

QY 181 FPAALPGQCMRFDMFQNDNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPSGANG 240
DB 181 FPAALPGQCMRFDMFQNDNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPSGANG 240

QY 241 GTGTPSTAPSGSGTSPGGSGCTSQKMAQCGGIGFSGCTTGVSGTTQKLNYYSQL 299
DB 241 GTGTPSTAPSGSGTSPGGSGCTSQKMAQCGGIGFSGCTTGVSGTTQKLNYYSQL 299

RESULT 4
US-09-230-222-1
Sequence 1, Application US/09230222A

1 GENERAL INFORMATION:
2 APPLICANT: CONVENTS, ANDRE C
3
4 APPLICANT: BUSCH, ALFRED
5
6 APPLICANT: BABCOCK, ANDRE C
7
8 TITLE OF INVENTION: DETERGENT COMPOSITION WITH HIGH ACTIVITY
9
10 TITLE OF INVENTION: CELLULOSE AND SOFTENING CLAYS
11
12 NUMBER OF SEQUENCES: 4
13
14 CORRESPONDENCE ADDRESS:
15
16 ADDRESSEE: THE PROCTER & GAMBLE COMPANY
17
18 STREET: 5299 SPRING GROVE AVENUE
19
20
21 CITY: CINCINNATI
22
23
24 STATE: OHIO
25
26 COUNTRY: USA
27
28 ZIP: 45217
29
30 COMPUTER READABLE FORM:
31
32 MEDIUM TYPE: Floppy disk
33
34 COMPUTER: IBM PC compatible
35

1 GENERAL INFORMATION:
2 APPLICANT: BAERK, ANDRE C.
3 APPLICANT: CEJEMANS, RAPHAEL ANGELINE A.
4 APPLICANT: BUSCH, ALFRED (NMN)
5 TITLE OF INVENTION: COMPACT DETERGENT COMPOSITIONS WITH
6 TITLE OF INVENTION: HIGH ACTIVITY CELLULOSE
7 NUMBER OF SEQUENCES: 4
8 CORRESPONDENCE ADDRESS:
9 ADDRESSEE: THE PROCTER & GAMBLE COMPANY
10 STREET: 11810 East Miami Road
11 CITY: CINCINNATI
12 STATE: OHIO
13 COUNTRY: USA
14 ZIP: 45253-8707
15 COMPUTER READABLE FORM:
16 MEDION TYPE: Diskette

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/081.328
FILING DATE: 11/19/93
CLASSIFICATION: 252
ATTORNEY/AGENT INFORMATION:
NAME: ZERRY, KIM WILLIAM
REGISTRATION NUMBER: 32,323
REFERENCE/DOCKET NUMBER: CM566
TELECOMMUNICATION INFORMATION:
TELEPHONE: 513-627-2885
TELEFAX: 513-627-0318
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-081-328-2

Query Match 69.9%; Score 1159; DB 1; Length 305;
Best Local Similarity 67.1%; Pred. No. 1.1e-86;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSPTVLTAAALPLVASAASGSGSTRYWDCKPCSCAMPKAAVOPVYACDANFOR 60
DB 1 MRSPTVLTAAALPLVASAASGSGSTRYWDCKPCSCAMPKAAVOPVYACDANFOR 58
QY 61 LSEFNVOGSCN-GGSAYSCADQTPMAVNDNLAYGFAATSIAGSSSSWCCACATLFTTSG 119
DB 59 ITDPAKSGCEPGEVAVASCADQTPMAVNDNLAYGFAATSIAGSSSSWCCACATLFTTSG 118
QY 120 PVAKTMVOSTSTGDLGSGNFDIAMPGGVGIENGCSQFGSLPGAQYGGISSRDQD 179
DB 119 PVAKTMVOSTSTGDLGSGNFDIAMPGGVGIENGCSQFGSLPGAQYGGISSRDQD 178
QY 180 SFAPALRGCOMFDMFONADNPTFTPOVCCPAEIVASGCKRNDSSP-VFTPPSG 238
DB 179 RFPDALRGCOMFDMFONADNPTFTPOVCCPAEIVASGCKRNDSSP-VFTPPSG 238
QY 239 NGGTGTPTSTAPSGGTS-----PGSGSGCTSOQMAOCGGIGFSGCTTCVSGTTCOKLN 292
DB 239 NGGTGTPTSTAPSGGTS-----PGSGSGCTSOQMAOCGGIGFSGCTTCVSGTTCOKLN 298
QY 293 DYVSQCL 299
DB 293 DYVSQCL 305

RESULT 7

US-08-232-249-2
Sequence 2, Application US/08232249
Patent No. 5610129
GENERAL INFORMATION:
APPLICANT: MCCORDUDALE, FINLAY (NMN)
APPLICANT: BUSCH, ALFRED (NMN)
TITLE OF INVENTION: DYE TRANSFER INHIBITING COMPOSITIONS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: THE PROCTER & GAMBLE COMPANY
STREET: 5299 SPRING GROVE AVENUE
CITY: CINCINNATI
STATE: OHIO
COUNTRY: USA
ZIP: 45217
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232.249
FILING DATE: 05-MAY-1994
CLASSIFICATION: 510
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 91202862.6
FILING DATE: 06-NOV-91
ATTORNEY/AGENT INFORMATION:
NAME: ALLEN, GEORGE W.
REGISTRATION NUMBER: 26,143
REFERENCE/DOCKET NUMBER: CM-395
TELECOMMUNICATION INFORMATION:
TELEPHONE: 513-627-5946
TELEFAX: 513-627-8118
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-232-249-2

Query Match 69.9%; Score 1159; DB 1; Length 305;
Best Local Similarity 67.1%; Pred. No. 1.1e-86;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSPTVLTAAALPLVASAASGSGSTRYWDCKPCSCAMPKAAVOPVYACDANFOR 60
DB 1 MRSPTVLTAAALPLVASAASGSGSTRYWDCKPCSCAMPKAAVOPVYACDANFOR 58
QY 61 LSEFNVOGSCN-GGSAYSCADQTPMAVNDNLAYGFAATSIAGSSSSWCCACATLFTTSG 119
DB 59 ITDPAKSGCEPGEVAVASCADQTPMAVNDNLAYGFAATSIAGSSSSWCCACATLFTTSG 118
QY 120 PVAKTMVOSTSTGDLGSGNFDIAMPGGVGIENGCSQFGSLPGAQYGGISSRDQD 179
DB 119 PVAKTMVOSTSTGDLGSGNFDIAMPGGVGIENGCSQFGSLPGAQYGGISSRDQD 178
QY 180 SFAPALRGCOMFDMFONADNPTFTPOVCCPAEIVASGCKRNDSSP-VFTPPSG 238
DB 179 RFPDALRGCOMFDMFONADNPTFTPOVCCPAEIVASGCKRNDSSP-VFTPPSG 236
QY 239 NGGTGTPTSTAPSGGTS-----PGSGSGCTSOQMAOCGGIGFSGCTTCVSGTTCOKLN 292
DB 239 NGGTGTPTSTAPSGGTS-----PGSGSGCTSOQMAOCGGIGFSGCTTCVSGTTCOKLN 298
QY 293 DYVSQCL 299
DB 293 DYVSQCL 305

RESULT 8

US-08-921-426-8
Sequence 8, Application US/08921426
Patent No. 5837847
GENERAL INFORMATION:
APPLICANT: ROYER, JOHN C
APPLICANT: MOYER, DONNA L
APPLICANT: YODER, WENDY T
APPLICANT: SHUSTER, JEFFREY R
TITLE OF INVENTION: NON-TOXIC, NON-TOXIGENIC, NON-PATHOGENIC
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: NO. 58378470 No. 5837847disk of No. 5837847th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/921,426
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/456,433
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: US 08/404,678
FILING DATE: 15-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Agis Dr., Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4216, 010-JS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-921-426-8

Query Match 69.9%; Score 1159; DB 2; Length 305;
Best Local Similarity 67.1%; Pred. No. 1,1e-86;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MESTPVLRTIAALALPVASAGSGGSTRYWDCKKPCSCAMPKAAVSQPVYACDANFOR 60
DB 1 MRSSPLPSAVVAALPVLAAL--DGRSTRYWDCKKPCSCAKAPVQCPFSCKNAFOR 58
QY 61 LSDPNVSGCN-GGSAYSCADQTPMAVNDLAVFAATSIAGSSSSWCCACVALFTTSG 119
DB 59 ITDPDAKSGCEPGVAVSCADQTPMAVNDLAVFAATSIAGSSSSWCCACVALFTTSG 118
QY 120 PVAAGTAVVQSTGTGDLGSGNFDIAMPGGVGIIFNGCSSQFGIPLAAYGGISRRQCD 179
DB 119 PVAAGTAVVQSTGTGDLGSGNFDIAMPGGVGIIFNGCSSQFGIPLAAYGGISRRQCD 178
QY 180 SFPAAPKRCGCMFDMFQKADNPFTFQVOCPAETIARSGCKRNDSDSP-VTPPSGG 238
DB 179 RFPDALRPGCTWRPDMFKADNPFSFRVQCPALVARTGRNDGDFPAVOIPSSST 238
QY 239 NGGTGTPSTAPSGGOTS-----PGGSGGCTSKMAQCGGIGFSGCTTCVSGTTCOKIN 292
DB 239 SSPVQPTSTSTSTSTSSPPVQPTTSGCTIARMAQCGGNGSGCTTCVAGSTCTKIN 298
QY 293 DYVSQCL 299
DB 299 DWYHQL 305

RESULT 9
US-08-833-642A-2
Sequence 2, Application US/08833642A
Patent No. 5883066
GENERAL INFORMATION:
APPLICANT: Ivan M. A. J. Herbots et al.
TITLE OF INVENTION: Liquid Detergent Compositions
CONTAINING CELLULOSE AND AMINE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jackie Ann Zurcher
ADDRESS: 255 E. Fifth Street
STREET: 1900 Chemed Center
CITY: Cincinnati
STATE: Ohio
COUNTRY: USA
ZIP: 45202
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/833,642A
FILING DATE: April 8, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Zurcher, J. A.
REGISTRATION NUMBER: P42,251
REFERENCE/DOCKET NUMBER: CM551C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (513) 977-8377
TELEFAX: (513) 977-8141
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-833-642A-2

Query Match 69.9%; Score 1159; DB 2; Length 305;
Best Local Similarity 67.1%; Pred. No. 1,1e-86;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MESTPVLRTIAALALPVASAGSGGSTRYWDCKKPCSCAMPKAAVSQPVYACDANFOR 60
DB 1 MRSSPLPSAVVAALPVLAAL--DGRSTRYWDCKKPCSCAKAPVQCPFSCKNAFOR 58
QY 61 LSDPNVSGCN-GGSAYSCADQTPMAVNDLAVFAATSIAGSSSSWCCACVALFTTSG 119
DB 59 ITDPDAKSGCEPGVAVSCADQTPMAVNDLAVFAATSIAGSSSSWCCACVALFTTSG 118
QY 120 PVAAGTAVVQSTGTGDLGSGNFDIAMPGGVGIIFNGCSSQFGIPLAAYGGISRRQCD 179
DB 119 PVAAGTAVVQSTGTGDLGSGNFDIAMPGGVGIIFNGCSSQFGIPLAAYGGISRRQCD 178
QY 180 SFPAAPKRCGCMFDMFQKADNPFTFQVOCPAETIARSGCKRNDSDSP-VTPPSGG 238
DB 179 RFPDALRPGCTWRPDMFKADNPFSFRVQCPALVARTGRNDGDFPAVOIPSSST 238
QY 239 NGGTGTPSTAPSGGOTS-----PGGSGGCTSKMAQCGGIGFSGCTTCVSGTTCOKIN 292
DB 239 SSPVQPTSTSTSTSTSSPPVQPTTSGCTIARMAQCGGNGSGCTTCVAGSTCTKIN 298
QY 293 DYVSQCL 299
DB 299 DWYHQL 305

RESULT 10
US-08-140-008A-4
Sequence 4, Application US/08140008A
Patent No. 5914306
GENERAL INFORMATION:
APPLICANT: SVENDSEN, Allan
APPLICANT: VON DER OSTEN, Claus
APPLICANT: CLAUSEN, Id Groth
APPLICANT: PATKAR, Shankant Anant
APPLICANT: BORCH, Kim
TITLE OF INVENTION: STABILIZED ENZYMES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5914306 No. 5914306disk of No. 5914306th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/140.008A
FILING DATE: 22-OCT-1993
CLASSIFICATION: 252
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3601.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 867-0123
TELEFAX: (212) 878-9655
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-140-008A-4

Query Match

69.3%; Score 1159; DB 2; Length 305;
Best Local Similarity 67.1%; Pred. No. 1.le-86;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSTPVLRITLAAALPLVSAASGSGSTRYMDCKPSCAMPKGAASQPVYACDANFOR 60
1 MRSSPLPSAVVAALPVTALAA--DGRSTRYMDCKPSCAMPKGAASQPVYACDANFOR 58
DB 61 LSDPVSQSGN-GGSAYSCADQTPMAVNDNLAYGPAATSIAGSSBSSWCCACVLTFTSG 119
59 ITDFKASGCBGEGVAVSCADQTPMAVNDNLAYGPAATSIAGSSBSSWCCACVLTFTSG 118
QY 120 PVAKTMYVOSTSTGDLGSGNPDIAFGGAGVGIENGSSQFGALPGAQYGGISSRDQD 179
119 PVAKKMYVOSTSTGDLGSGNPDIAFGGAGVGIENGSSQFGALPGAQYGGISSRDQD 178
DB 180 SFPAFLPGCCMRPFMDNADNPTFTFOVOCPAEIVARSCKRRNDSSP-VFTPPSGG 238
179 RFPDAKPGCYWRPFMDNADNPTFTFOVOCPAEIVARSCKRRNDSSP-VFTPPSGG 238
QY 239 NGGTGPTSTAPSGQTS-----PGGSGCTSQKMAQCGGIGFSGCTTCVSGTTQCKLN 292
DB 239 SSPVNGPTSTSTSTSTSSPPVQPTTPSGCTAARMACGGAGWGGCTTCVSGTTQCKLN 298
QY 293 DYVSQCL 299
DB 299 DMVHQL 305

RESULT 11

US-08-836-340-1
Sequence 1, Application US/08836340
Patent No. 5916798
GENERAL INFORMATION:
APPLICANT: Lund, Henrik
APPLICANT: Pedersen, Hanne H st
TITLE OF INVENTION: A Method of Obtaining a Cellulosic Textile
TITLE OF INVENTION: Fabric with Reduced Tendency to Pilling Formation
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59167980 No. 5916798disk of No. 5916798th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/836,340
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: *Humicola insolens*
STRAIN: DSM 1800
US-08-836-340-1

Query Match

69.3%; Score 1159; DB 2; Length 305;
Best Local Similarity 67.1%; Pred. No. 1.le-86;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSTPVLRITLAAALPLVSAASGSGSTRYMDCKPSCAMPKGAASQPVYACDANFOR 60
1 MRSSPLPSAVVAALPVTALAA--DGRSTRYMDCKPSCAMPKGAASQPVYACDANFOR 58
DB 61 LSDPVSQSGN-GGSAYSCADQTPMAVNDNLAYGPAATSIAGSSBSSWCCACVLTFTSG 119
59 ITDFKASGCBGEGVAVSCADQTPMAVNDNLAYGPAATSIAGSSBSSWCCACVLTFTSG 118
QY 120 PVAKTMYVOSTSTGDLGSGNPDIAFGGAGVGIENGSSQFGALPGAQYGGISSRDQD 179
119 PVAKKMYVOSTSTGDLGSGNPDIAFGGAGVGIENGSSQFGALPGAQYGGISSRDQD 178
DB 180 SFPAFLPGCCMRPFMDNADNPTFTFOVOCPAEIVARSCKRRNDSSP-VFTPPSGG 238
179 RFPDAKPGCYWRPFMDNADNPTFTFOVOCPAEIVARSCKRRNDSSP-VFTPPSGG 238
QY 239 NGGTGPTSTAPSGQTS-----PGGSGCTSQKMAQCGGIGFSGCTTCVSGTTQCKLN 292
DB 239 SSPVNGPTSTSTSTSTSSPPVQPTTPSGCTAARMACGGAGWGGCTTCVSGTTQCKLN 298
QY 293 DYVSQCL 299
DB 299 DMVHQL 305

RESULT 12

US-08-389-423-2
Sequence 2, Application US/08389423
Patent No. 5948672
GENERAL INFORMATION:
APPLICANT: Rasmussen, Grethe
APPLICANT: Mikkelson, Jan Moller
APPLICANT: Schulten, Martin
APPLICANT: Patkar, Shankar A.
TITLE OF INVENTION: A Cellulase Preparation Comprising an
TITLE OF INVENTION: Endoglucanase Enzyme
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59486720 No. 5948672disk of No. 5948672th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/389,423
FILING DATE: 14-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3469.214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-389-423-2

Query Match 69.9%; Score 1159; DB 2; Length 305;
Best Local Similarity 67.1%; Pred. No. 1.1e-86;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSTPVLRTTLAALPLVSAASGSGSTRYWDCKPSCAMPGKAAYSPVACDANFOR 60
DB 1 MRSFPLPSAVVAALPVLAAL--DGRSTRYWDCKPSCGMAKAPVAVQVESCANFOR 58
QY 61 LSDPNVOSGN-GSAYSCADQTPMAVNDNLAVGPAATSIAGSSHSNCCACVATLFTSG 119
DB 59 ITDFPAKSGCEPGVAVSCADQTPMAVNDNLAVGPAATSIAGSNEMAGCCACVATLFTSG 118
QY 123 PVAGKTNVOSTSTGDLGSGNOFDIAMPGGVGIENGCSQFGGLPQAGYGGISSRQCD 179
DB 119 PVAGKTNVOSTSTGDLGSGNHFDLNIIPGGVGIENGCTPQFGGLPQAGYGGISSRQCD 178
QY 180 SEPAPLPGGQMRPDMFQNAADNPTFPQVQCPAEIVASGCKRNDSSFP-VETPPSGG 238
DB 179 RFPDAKPGCYMRPDMFQNAADNPTFPQVQCPAEIVATGCKRNDGNFPAVQ-ESSST 238
QY 239 NKGITPTSTIAPSGQTS-----PGSGSGCTSQKMACGGIGSGCTTCVSGTTCQKLN 292
DB 239 SSPVQPTSTSTSTSTSTSSFPVQPTTSGCTAERMAQCGGNGSGCTTCVAGSTCTKLN 298
QY 293 DYVSQCL 299
DB 299 DWTQCL 305

RESULT 13

US-08-816-915-8
Sequence 8, Application US/08816915
Patent No. 6060305
GENERAL INFORMATION:
APPLICANT: Royer, John C.
APPLICANT: Moyer, Donna L.
APPLICANT: Yoder, Wendy J.
APPLICANT: Stuster, Jeffrey R.
TITLE OF INVENTION: NON-TOXIC, NON-TOXIGENIC, NON-PATHOGENIC
TITLE OF INVENTION: FUSARIM EXPRESSION SYSTEM
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6060305 No. 6060305disk of No. 6060305th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/816,915
FILING DATE: 13-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Agtis Dr., Cheryl H.

REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4216.240-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-816-915-8

Query Match 69.9%; Score 1159; DB 3; Length 305;
Best Local Similarity 67.1%; Pred. No. 1.1e-86;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSTPVLRTTLAALPLVSAASGSGSTRYWDCKPSCAMPGKAAYSPVACDANFOR 60
DB 1 MRSFPLPSAVVAALPVLAAL--DGRSTRYWDCKPSCGMAKAPVAVQVESCANFOR 58
QY 61 LSDPNVOSGN-GSAYSCADQTPMAVNDNLAVGPAATSIAGSSHSNCCACVATLFTSG 119
DB 59 ITDFPAKSGCEPGVAVSCADQTPMAVNDNLAVGPAATSIAGSNEMAGCCACVATLFTSG 118
QY 120 PVAGKTNVOSTSTGDLGSGNOFDIAMPGGVGIENGCSQFGGLPQAGYGGISSRQCD 179
DB 119 PVAGKTNVOSTSTGDLGSGNHFDLNIIPGGVGIENGCTPQFGGLPQAGYGGISSRQCD 178
QY 180 SEPAPLPGGQMRPDMFQNAADNPTFPQVQCPAEIVASGCKRNDSSFP-VETPPSGG 238
DB 179 RFPDAKPGCYMRPDMFQNAADNPTFPQVQCPAEIVATGCKRNDGNFPAVQ-ESSST 238
QY 239 NKGITPTSTIAPSGQTS-----PGSGSGCTSQKMACGGIGSGCTTCVSGTTCQKLN 292
DB 239 SSPVQPTSTSTSTSTSTSSFPVQPTTSGCTAERMAQCGGNGSGCTTCVAGSTCTKLN 298
QY 293 DYVSQCL 299
DB 299 DWTQCL 305

RESULT 14

US-09-189-060B-56
Sequence 56, Application US/09189060B
Patent No. 6270968
GENERAL INFORMATION:
APPLICANT: Dalboge, Henrik
APPLICANT: Sandal, Thomas
APPLICANT: Kaupinen, Markus
APPLICANT: Borge, Diderichsen
TITLE OF INVENTION: Method of Providing No. 6270968e1 DNA Sequences
FILE REFERENCE: 4772.204-US
CURRENT APPLICATION NUMBER: US/09/189,360B
CURRENT FILING DATE: 1998-11-10
PRIOR APPLICATION NUMBER: PCT/DK97/00216
PRIOR FILING DATE: 1997-05-12
NUMBER OF SEQ ID NOS: 74
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 56
LENGTH: 305
TYPE: PRT
ORGANISM: Humicola insolens
FEATURE:
NAME/KEY: SIGNAL
LOCATION: (1)...(21)
US-09-189-060B-56

Query Match 69.9%; Score 1159; DB 3; Length 305;
Best Local Similarity 67.1%; Pred. No. 1.1e-86;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSTPVLRTTLAALPLVSAASGSGSTRYWDCKPSCAMPGKAAYSPVACDANFOR 60

Search completed: July 7, 2004, 08:46:47
Job time : 23 secs

```
Db 1 MRSSPILP2SAVVAALPVIALAA--DGRSTRFWDCKKSCGMAKKAPVNPVFSCKNPFQR 58
Qy 61 ISDPNVQSGCN-GGSAYSACADQTPMAVNDNLAYGPAATSIAGSSSSCCACVLTFTSG 119
Db 59 ITDFDAKSGCBPGVAVYSCADQTPMAVNDNDFALGPAATSIAGSNEAGWCCACVLTFTSG 118
Qy 120 PVAGKTMVWOSTSTGGDLGASNOPDILAMPGGGYSIFNGCSQFGLPGAQYGISRPDQD 179
Db 119 PVAGKXVWOSTSTGGDLGASNFDLNIPEGGVGLFDGCTPFGGLPGQRYGISRNECD 178
Qy 180 SFPAPLKPGCCWRPFQWPNADNPTFTFOVQCPAEIVARSGCKRNDSSFP-VTPPSGG 238
Db 179 RFPDAKPGCYRPFQWPNADNPSFSFRQVQCPAEIVARIGCRNDGNFPAVQIPSSST 238
Qy 239 NGGTGTPSTAPGSGQTS-----PGGSGCTSQMAQCGGIGFSCTTCVSGTTCCKIN 292
Db 239 SSPVNPSTSTSTSTSSPPVQPTTPSGCTABRMWQCGNGMSGCTTCVAGSTCTKIN 298
Qy 293 DYSQCL 299
Db 299 DMYHQL 305
```

RESULT 15

```
US-09-230-665-2
; Sequence 2, Application US/09230665
; Patent No. 6322595
; GENERAL INFORMATION:
; APPLICANT: Boyer, Stanton I
; TITLE OF INVENTION: Detergent Composition Comprising Two Cellulase
; TITLE OF INVENTION: Components, with and without a Cellulose-Binding Domain
; FILE REFERENCE: 6191 Sequence listing (8 Sequences)
; Patent No. 6322595
; CURRENT APPLICATION NUMBER: US/09/230,665
; CURRENT FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: 60/023,125
; EARLIER FILING DATE: 1996-07-30
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Humicola insolens
US-09-230-665-2
```

Query Match

69.9%; Score 1159; DB 4; Length 305;

Best Local Similarity 67.1%; Pred. No. 1,1e-86;

Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

```
Qy 1 MRSTPVLRTTLAALPIVAAASGSGSTRYWDCCCKPSCAMPGKAIVSOPVYACDANFOR 60
Db 1 MRSSPILP2SAVVAALPVIALAA--DGRSTRFWDCKKSCGMAKKAPVNPVFSCKNPFQR 58
Qy 61 LSDNVQSGCN-GGSAYSACADQTPMAVNDNLAYGPAATSIAGSSSSWCCACVLTFTSG 119
Db 59 ITDFDAKSGCBPGVAVYSCADQTPMAVNDNDFALGPAATSIAGSNEAGWCCACVLTFTSG 118
Qy 120 PVAGKTMVWOSTSTGGDLGASNOPDILAMPGGGYSIFNGCSQFGLPGAQYGISRPDQD 179
Db 119 PVAGKXVWOSTSTGGDLGASNFDLNIPEGGVGLFDGCTPFGGLPGQRYGISRNECD 178
Qy 180 SFPAPLKPGCCWRPFQWPNADNPTFTFOVQCPAEIVARSGCKRNDSSFP-VTPPSGG 238
Db 179 RFPDAKPGCYRPFQWPNADNPSFSFRQVQCPAEIVARIGCRNDGNFPAVQIPSSST 238
Qy 239 NGGTGTPSTAPGSGQTS-----PGGSGCTSQMAQCGGIGFSCTTCVSGTTCCKIN 292
Db 239 SSPVNPSTSTSTSSPPVQPTTPSGCTABRMWQCGNGMSGCTTCVAGSTCTKIN 298
Qy 293 DYSQCL 299
Db 299 DMYHQL 305
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 08:46:10 ; Search time 51 Seconds
(without alignments)
1824.978 Million cell updates/sec

Title: US-10-007-521-12

Perfect score: 1659

Sequence: 1 MRSPTVKTITAAALPIVAS.....TTCVSGCTCKKNDYISQCL 299

Scoring table: BLOSUM62

Gapop 10.0, Capext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_5/prodata/1/pubpa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/1/pubpa/PCF_NEW_PUB.pep:*
3: /cgn2_6/prodata/1/pubpa/US06_NEW_PUB.pep:*
4: /cgn2_6/prodata/1/pubpa/US06_PUBCOMB.pep:*
5: /cgn2_6/prodata/1/pubpa/US07_NEW_PUB.pep:*
6: /cgn2_6/prodata/1/pubpa/PCFUS_PUBCOMB.pep:*
7: /cgn2_6/prodata/1/pubpa/US08_NEW_PUB.pep:*
8: /cgn2_6/prodata/1/pubpa/US08_PUBCOMB.pep:*
9: /cgn2_6/prodata/1/pubpa/US09A_PUBCOMB.pep:*
10: /cgn2_6/prodata/1/pubpa/US09B_PUBCOMB.pep:*
11: /cgn2_6/prodata/1/pubpa/US09C_PUBCOMB.pep:*
12: /cgn2_6/prodata/1/pubpa/US09_NEW_PUB.pep:*
13: /cgn2_6/prodata/1/pubpa/US10A_PUBCOMB.pep:*
14: /cgn2_6/prodata/1/pubpa/US10B_PUBCOMB.pep:*
15: /cgn2_6/prodata/1/pubpa/US10C_PUBCOMB.pep:*
16: /cgn2_6/prodata/1/pubpa/US10_NEW_PUB.pep:*
17: /cgn2_6/prodata/1/pubpa/US60_NEW_PUB.pep:*
18: /cgn2_6/prodata/1/pubpa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1659	100.0	299	14	US-10-007-521-12
2	1216	73.3	296	15	US-10-369-493-4204
3	1159	69.9	305	9	US-09-735-787-2
4	1159	69.9	305	14	US-10-138-870-2
5	1122	67.6	201	10	US-09-261-329-5
6	1039	62.6	235	8	US-08-841-636A-31
7	1021	61.5	308	14	US-10-007-521-6
8	994.5	59.9	297	14	US-10-007-521-4
9	962	58.0	310	14	US-10-007-521-22
10	951	57.3	295	14	US-10-007-521-8
11	939.5	56.6	376	9	US-09-735-787-4
12	939.5	56.6	376	14	US-10-138-870-4
13	938	55.7	201	10	US-09-261-329-4
14	923.5	55.7	294	14	US-10-007-521-24
15	910.5	54.9	349	14	US-10-007-521-10

16	887.5	53.5	202	10	US-09-261-329-1
17	887	53.5	298	14	US-10-007-521-18
18	884	53.3	293	14	US-10-007-521-20
19	819	49.4	203	10	US-09-261-329-6
20	801.5	48.3	223	12	US-10-432-290-38
21	796.5	47.8	225	14	US-10-007-521-2
22	792.5	47.8	228	12	US-10-432-290-24
23	777	46.8	205	10	US-09-261-329-7
24	769.5	46.4	338	12	US-10-416-328-4
25	769.5	46.4	338	12	US-10-432-290-7
26	769.5	46.4	387	12	US-10-416-328-5
27	769.5	46.4	387	12	US-10-432-290-9
28	766.5	46.2	245	12	US-10-432-290-29
29	757.5	45.7	366	12	US-10-416-328-2
30	757.5	45.7	366	12	US-10-432-290-3
31	756.5	45.6	338	12	US-10-416-328-1
32	756.5	45.6	338	12	US-10-432-290-1
33	752.5	45.4	202	10	US-09-261-329-3
34	745	44.9	360	12	US-10-416-328-3
35	745	44.9	360	12	US-10-432-290-5
36	725	43.7	222	14	US-10-007-521-14
37	722.5	43.6	202	10	US-09-261-329-2
38	718	43.3	203	10	US-09-261-329-9
39	688.5	41.5	226	14	US-10-007-521-16
40	686.5	41.4	346	12	US-10-416-328-6
41	686.5	41.4	346	12	US-10-432-290-11
42	672	40.5	203	10	US-09-261-329-8
43	587	35.4	138	14	US-10-007-521-26
44	496	29.9	235	10	US-09-261-329-10
45	419	25.3	211	10	US-09-261-329-11

ALIGNMENTS

RESULT 1
US-10-007-521-12
Sequence 12, Application US/10007521
Publication No. US20030054539A1
GENERAL INFORMATION:
APPLICANT: Schulte, Martin
Andersen, Lene N.
Lassen, Soren F.
Kauppinen, Markus S.
Lange, Lene
Nielsen, Ruby I.
Ihara, Michiko
Takagi, Shinobu
TITLE OF INVENTION: No. US20030054539A1 Endoglucaes
NUMBER OF SEQUENCES: 109
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. US20030054539A1 No. US20030054539A1disk of No. US2003001
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/007,521
FILING DATE: 10-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/551,136
FILING DATE: 21-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4366,200-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 299 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-007-521-12

Query Match 100.0%; Score 1659; DB 14; Length 299;
Best Local Similarity 100.0%; Pred. No. 1,9e-131;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSTPVLRTTLAALPLVASAASGSGSTRYMDCCPKSCAMPKGAASQPVYACDANFOR 60
DB 1 MSTPVLRTTLAALPLVASAASGSGSTRYMDCCPKSCAMPKGAASQPVYACDANFOR 60
QY 61 LSDPVSQGCNGSANSQADQTPMAVNDNLAYGFAATSIAGSSSSMCACVALPTSGP 120
DB 61 LSDPVSQGCNGSANSQADQTPMAVNDNLAYGFAATSIAGSSSSMCACVALPTSGP 120
QY 121 VAGKTMVOSTSTGDLGSGNCPDIAMPQGGVGFNGCSSQFGSLPGAQYGGISSRDQDS 180
DB 121 VAGKTMVOSTSTGDLGSGNCPDIAMPQGGVGFNGCSSQFGSLPGAQYGGISSRDQDS 180
QY 181 PAPLPKPGCCQREDFWQNDNPTFTFOVQCPAIVARSGCKRNDSSPFVFTPPSGNG 240
DB 181 PAPLPKPGCCQREDFWQNDNPTFTFOVQCPAIVARSGCKRNDSSPFVFTPPSGNG 240
QY 241 GGTGTSTAPSGSGTSPGSGSGCTSOXMAQCGIGFSGCTTCVSGTTCQKLANDYISQL 299
DB 241 GGTGTSTAPSGSGTSPGSGSGCTSOXMAQCGIGFSGCTTCVSGTTCQKLANDYISQL 299

RESULT 2

US-10-369-493-4204
Sequence 4204, Application US/10369493
Publication No. US2003023675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xiandeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIOR FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 4204
LENGTH: 296
TYPE: PRT
ORGANISM: Neurospora crassa
FEATURES:
NAME/KEY: unsure
LOCATION: (1)-(296)
OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-4204

Query Match 73.3%; Score 1216; DB 15; Length 296;
Best Local Similarity 72.2%; Pred. No. 3.5e-94;
Matches 215; Conservative 27; Mismatches 36; Indels 24; Gaps 3;

QY 15 LPLVSAASGSGSTRYMDCCPKSCAMPKGAASQPVYACDANFORLSDPVSQGCNGS 74
DB 1 LPLVSAASGSGSTRYMDCCPKSCAMPKGAASQPVYACDANFORLSDPVSQGCNGS 60

QY 75 AYSCADQTPMAVNDNLAYGFAATSIAGSSSSMCACV-----ALTE 116
DB 61 AYTCANNSPMAVNDNLAYGFAATSIAGSSSSMCACVAYSLQLLGITRLTETALTF 120
QY 117 TSGPVAQKTMVOSTSTGDLGSGNCPDIAMPQGGVGFNGCSSQFGSLPGAQYGGISSRD 176
DB 121 TSGPVAQKTMVOSTSTGDLGSGNCPDIAMPQGGVGFNGCSSQFGSLPGAQYGGISSRD 180
QY 177 QCSFPALKPGCCQREDFWQNDNPTFTFOVQCPAIVARSGCKRNDSSPFVFTPPS 236
DB 181 QCSFPALKPGCCQREDFWQNDNPTFTFOVQCPAIVARSGCKRNDSSPFVFTPPS 240
QY 237 GANSGTGTSTAPSGSGTSPGSGSGCTSOXMAQCGIGFSGCTTCVSGTTCQKLANDYIS 296
DB 241 GGSNSSTPT-TPSS-----GGSGCTADKXKAGGSGMSGCTNPGSGSTCKTLNDYH 294
QY 297 QC 298
DB 295 QC 296

RESULT 3

US-09-735-787-2
Sequence 2, Application US/09735787
Patent No. US20010036910A1
GENERAL INFORMATION:
APPLICANT: Rasmussen, Grethe
Mikkelsen, Jan Koller
Schulein, Martin
Packer, Shankant A.
Hagen, Fred
TITLE OF INVENTION: A Cellulase Preparation Comprising an
Endoglucanase Enzyme
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. US20010036910A10 No. US20010036910A1disk of No. US200100:
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/735,787
FILING DATE: 13-Dec-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/189,028
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Lambitis, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3469,214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-735-787-2

Query Match 69.9%; Score 1159; DB 9; Length 305;
Best Local Similarity 67.1%; Pred. No. 2.3e-89;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

```

Qy 1 MSSTPVLTITLAAALPLVNASASGSGSTRYWDCKPSCAMPKRAVSPVYACDANFOR 60
Db 1 MSSSPLPSAVVAALPLVLAAL--DGRSTRYWDCKPSCGMAKKAPVNOVPSCKANFOR 58
Qy 61 LSPFNVOGSCN-GGSAYSCADQTPMAVNDMLAYGPAATSIAGSSESSWCCACVATLFTSG 119
Db 59 ITDFPAKSGCEPGGVAYSCADQTPMAVNDPLAGPAATSIAGSSESSWCCACVATLFTSG 118
Qy 120 PVAKTMYVOSTSTGDLGSGNOFDIAMEGSGVGIENGSGSOPGALPGAQYGISRRDCCD 179
Db 119 PVAKKMYVOSTSTGDLGSGNHFDLNIPEGSGVGIENGSGSOPGALPGAQYGISRRNCD 178
Qy 180 SPFAPLKPCCWRPFQMANPCTTPOVQCCPAEIVASGCKRNDSSFP-VFTPPSGG 238
Db 179 RFPDLKPGCYWRPFQMANPCTTPOVQCCPAEIVASGCKRNDSSFP-VFTPPSGG 238
Qy 239 NGCTGPTSTAPSGQTS-----PGSGSGCTSQKMAQCGGIGFSGCTTCVSGTTCOKLN 292
Db 239 SSPVNGPTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGGKMSGCTTCVAGSTCTKIN 298
Qy 293 DYISQCL 299
Db 299 DWYHQCL 305

```

RESULT 4

```

US-10-138-870-2
; Sequence 2, Application US/10138870
; Publication No. US20030119167A1

```

GENERAL INFORMATION:

```

APPLICANT: Rasmussen, Gretche
           Mikelsen, Jan Moller
           Schuelein, Martin
           Packar, Shankant A.
           Hagen, Fred

```

```

TITLE OF INVENTION: A Cellulase Preparation Comprising an
                   Endoglucanase Enzyme

```

```

NUMBER OF SEQUENCES: 33

```

```

CORRESPONDENCE ADDRESS:

```

```

ADDRESS: No. US20030119167A10 No. US20030119167A1disk of No. US200301191
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/138,870
FILING DATE: 03-May-2002
CLASSIFICATION: <Unknown>

```

```

PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: US/09/735,787
FILING DATE: 13-Dec-2000

```

```

APPLICATION NUMBER: 09/189,028
FILING DATE: <Unknown>

```

```

ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.

```

```

REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3469,214-US

```

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655

```

```

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

```

```

LENGTH: 305 amino acids

```

```

TYPE: amino acid

```

```

MOLECULAR TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

```

US-10-138-870-2

```

```

Query Match 69.9%; Score 1159; DB 14; Length 305;
Best Local Similarity 67.1%; Pred. No. 2,3e-89;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

```

```

Qy 1 MSSTPVLTITLAAALPLVNASASGSGSTRYWDCKPSCAMPKRAVSPVYACDANFOR 60
Db 1 MSSSPLPSAVVAALPLVLAAL--DGRSTRYWDCKPSCGMAKKAPVNOVPSCKANFOR 58
Qy 61 LSPFNVOGSCN-GGSAYSCADQTPMAVNDMLAYGPAATSIAGSSESSWCCACVATLFTSG 119
Db 59 ITDFPAKSGCEPGGVAYSCADQTPMAVNDPLAGPAATSIAGSSESSWCCACVATLFTSG 118
Qy 120 PVAKTMYVOSTSTGDLGSGNOFDIAMEGSGVGIENGSGSOPGALPGAQYGISRRDCCD 179
Db 119 PVAKKMYVOSTSTGDLGSGNHFDLNIPEGSGVGIENGSGSOPGALPGAQYGISRRNCD 178
Qy 180 SPFAPLKPCCWRPFQMANPCTTPOVQCCPAEIVASGCKRNDSSFP-VFTPPSGG 238
Db 179 RFPDLKPGCYWRPFQMANPCTTPOVQCCPAEIVASGCKRNDSSFP-VFTPPSGG 238
Qy 239 NGCTGPTSTAPSGQTS-----PGSGSGCTSQKMAQCGGIGFSGCTTCVSGTTCOKLN 292
Db 239 SSPVNGPTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGGKMSGCTTCVAGSTCTKIN 298
Qy 293 DYISQCL 299
Db 299 DWYHQCL 305

```

RESULT 5

```

US-09-261-329-5
; Sequence 5, Application US/09261329
; Publication No. US20030092097A1

```

GENERAL INFORMATION:

```

APPLICANT: Andersen, Kim
           Schuelein, Martin
           Christiansen, Lars

```

```

APPLICANT: Damgaard, Bo

```

```

APPLICANT: Von Der Osten, Claus

```

```

TITLE OF INVENTION: Cellulase Variants

```

```

FILE REFERENCE: 4887,204-US

```

```

CURRENT APPLICATION NUMBER: US/09/261,329

```

```

CURRENT FILING DATE: 1999-03-03

```

```

EARLIER FILING DATE: 1996-09-17

```

```

NUMBER OF SEQ ID NOS: 26

```

```

SOFTWARE: FastSeq for Windows Version 3.0

```

```

SEQ ID NO 5

```

```

TYPE: PRT

```

```

ORGANISM: Cellulase variants

```

```

US-09-261-329-5

```

```

Query Match 67.6%; Score 1122; DB 10; Length 201;
Best Local Similarity 100.0%; Pred. No. 1,8e-86;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 24 GSGGSTRYWDCKPSCAMPKRAVSPVYACDANFORLSDTVNVOGSGGSAVSCADQTP 83
Db 1 GSGGSTRYWDCKPSCAMPKRAVSPVYACDANFORLSDTVNVOGSGGSAVSCADQTP 60
Qy 84 NAYVNDMLAYGPAATSIAGSSESSWCCACVATLFTSGPVAGKMYVOSTSTGDLGSGNOFD 143
Db 61 NAYVNDMLAYGPAATSIAGSSESSWCCACVATLFTSGPVAGKMYVOSTSTGDLGSGNOFD 120
Qy 144 IMPGGGVGIFNGCSGQFGLPGAQYGISRRDCCSPFAPLKPCCWRPFQMANPCTT 203
Db 121 IMPGGGVGIFNGCSGQFGLPGAQYGISRRDCCSPFAPLKPCCWRPFQMANPCTT 180
Qy 204 PTFQVQCCPAEIVASGCKR 223

```

Db 181 FTQOVQCPAEIVASGCKR 200

RESULT 6
US-08-841-636A-31
Sequence 31, Application US/08841636A
Publication No. US20020168751A1
GENERAL INFORMATION:
APPLICANT: Miettinen-Oinonen, Arja
APPLICANT: Londeborough, John
APPLICANT: Vehmanen, Jari
APPLICANT: Haakana, Hei
APPLICANT: Mätyl, Arja
APPLICANT: Lantto, Raija
APPLICANT: Elvainen, Minna
APPLICANT: Joutsen, Vesa
APPLICANT: Palohelmo, Marja
APPLICANT: Suominen, Pirkko
TITLE OF INVENTION: NOVEL CELLULASES, THE GENES ENCODING THEM AND
TITLE OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/841,636A
FILING DATE: 30-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/005,335
FILING DATE: 17-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/007,926
FILING DATE: 04-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/020,840
FILING DATE: 28-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/732,181
FILING DATE: 16-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: ECT/F196/00550
FILING DATE: 17-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Timothy J. Shea, Jr.
REGISTRATION NUMBER: 41,306
REFERENCE/DOCKET NUMBER: 1716, 0510005/MAC/TJS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)371-2600
TELEFAX: (202)371-2540
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Melanocarpus albomyces
STRAIN: ALKO4237
FEATURE:
NAME/KEY: Protein
LOCATION: 1..235
OTHER INFORMATION: /label= 20K-cellulase

US-08-841-636A-31

Query Match 62.6%; Score 1039; DB 8; Length 235;
Best Local Similarity 76.3%; Pred. No. 2,1e-79;
Matches 180; Conservative 27; Mismatches 27; Indels 2; Gaps 1;

QY 1 NASTVLELTLLAALPLVNASAGSGSTYRWCKCPKAPGAANVSQYVACANCR 60
DB 1 KRSTVLELTLLAALPLGALAA--NGQSTYRWCKCPKAPGAANVSQYVACANCR 58
QY 61 LQDENVSGCGGSAVSCADOTPAVNDNLAYFAATSIAGSSSSMCCCYALFTSGP 120
DB 59 IHPDPAVSGCBGPAPSCADHSPVHINDNLSTGFATLISQTEBSMCCCYALFTSGP 118
QY 121 VAGKTMVQSTSTGDLGSGNQFDLAMPGSGVGIENGCSGSGFGLPGAQYGGISSRDQDS 180
DB 119 VAGKTMVQSTSTGDLGSGNHFDLHPGGVGLFDGCTPQFGGLPGAQYGGISSRDQDS 178
QY 181 FPAELKPGCCQRPWFQNDNPTFTFQOVQCPAEIVASGCKRNDSSPFVTPPS 236
DB 179 FPEPLKPGCCQRPWFQNDNPSFTFERVQCPSELVARTGCKRHDDGGFAVFPAPS 234

RESULT 7

US-10-007-521-6
Sequence 6, Application US/10007521
Publication No. US20030054539A1
GENERAL INFORMATION:
APPLICANT: Schuelein, Martin
APPLICANT: Andersen, Lene N.
APPLICANT: Kauppinen, Markus S.
APPLICANT: Lange, Lene
APPLICANT: Nielsen, Ruby I.
APPLICANT: Ihara, Michiko
APPLICANT: Takagi, Shinobu
TITLE OF INVENTION: No. US20030054539A1el Endoglucanases
NUMBER OF SEQUENCES: 109
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. US20030054539A1o No. US20030054539A1disk of No. US200300
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/007,521
FILING DATE: 10-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/651,136
FILING DATE: 21-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Lambitis, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4366, 200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 308 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-007-521-6

Query Match: 61.5%; Score 1021; DB 14; Length 308;
Best Local Similarity 60.9%; Pred. No. 9.4e-78;
Matches 185; Conservative 43; Mismatches 62; Indels 14; Gaps 7;

QY 9 TTAALPLVA-SAASGSGSTRYWDCCPSGAMPKGAASQPYACANFORLSD-FNV 66
DB 6 TTGFLALPYALDQISGIGQTRRYWDCCPSGAMPKGP-SSPYQACDKNPNLMDGST 64
QY 67 QGCGN-GGSAVSCADTTPWAVNDNLATGPAATISAGSSBSWCACTALTFTSGPVAKT 125
DB 65 RSGCGAGSAYMCSQSPWAVSDLSYGMVAAYLAGSSBQWCCACELTFTSGPVAKK 124
QY 126 MVVOSTSTGDLGSDNFDIAMPGGGSGVIFNGCSSQFGALP---GAQYGGISRDQCSFP 182
DB 125 MIVQATNTGDLGDNHFDLALPGGSGVIFNACTDQYGAPEMGWGRYGGIHSKECESFP 184
QY 183 APLKPGQMRPFQNDNPTFTTFOVQCPAEIVASGCKNDSDSFP-VTPPSGANGG 241
DB 185 EALKPGQMRPFQNDNPTFTTFOVQCPAEIVASGCKNDSDSFP-VTPPSGANGG 244
QY 242 TGTPTSTARGSGQTS-----PGGSGCTGQNAQCGGIGSGCTTCVSGTTCCQKLNXY 295
DB 245 VNQPTSTSTSTSTSTSPVQPTTPTSGCTAERWACCGGNGSGCTTCVAGSTCTKINDMY 304
QY 296 SQCL 299
DB 305 HQCL 308

RESULT 8
US-10-007-521-4

? Sequence 4, Application US/10007521
? Publication No. US20030054539A1
? GENERAL INFORMATION:
? APPLICANT: Schulein, Martin
? Andersen, Lene N.
? Lassen, Soren F.
? Kauppinen, Markus S.
? Lange, Lene
? Nielsen, Ruby I.
? Ihara, Michiko
? Takagi, Shinobu
? TITLE OF INVENTION: No. US20030054539A1el Endoglucanases
? NUMBER OF SEQUENCES: 109
? CORRESPONDENCE ADDRESS:
? ADDRESS: No. US20030054539A1o No. US20030054539A1disk of No. US20030054539A1
? STREET: 405 Lexington Avenue, 64th Floor
? CITY: New York
? STATE: New York
? COUNTRY: United States of America
? ZIP: 10174-6401
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/10/007,521
? FILING DATE: 10-Dec-2001
? CLASSIFICATION: <Unknown>
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/651,136
? FILING DATE: 21-MAY-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Lambiris, Elias J.
? REGISTRATION NUMBER: 33,728
? REFERENCE/DOCKET NUMBER: 4366.200-US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212-867-0123
? TELEFAX: 212-878-9655
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 297 amino acids

? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULAR TYPE: protein
? SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-007-521-4

Query Match: 59.9%; Score 994.5; DB 14; Length 297;
Best Local Similarity 60.7%; Pred. No. 1.5e-75;
Matches 181; Conservative 45; Mismatches 59; Indels 13; Gaps 8;

QY 9 TTAALPLVA-SAASGSGSTRYWDCCPSGAMPKGAASQPYACANFORLSD-FNV 66
DB 6 TTGFLALPYALDQISGIGQTRRYWDCCPSGAMPKGP-SSPYQACDKNPNLMDGST 64
QY 67 QGCGN-GGSAVSCADTTPWAVNDNLATGPAATISAGSSBSWCACTALTFTSGPVAKT 125
DB 65 RSGCGAGSAYMCSQSPWAVSDLSYGMVAAYLAGSSBQWCCACELTFTSGPVAKK 124
QY 126 MVVOSTSTGDLGSDNFDIAMPGGGSGVIFNGCSSQFGALP---GAQYGGISRDQCSFP 182
DB 125 MIVQATNTGDLGDNHFDLALPGGSGVIFNACTDQYGAPEMGWGRYGGIHSKECESFP 184
QY 183 APLKPGQMRPFQNDNPTFTTFOVQCPAEIVASGCKR-NDSSFFVFTPPSGANGG 241
DB 185 EALKPGQMRPFQNDNPTFTTFOVQCPAEIVASGCKR-NDSSFFVFTPPSGANGG 242
QY 242 TGTPTSTARGSGQTSRGGSGCTGQNAQCGGIGSGCTTCVSGTTCCQKLNXY 299
DB 243 TSTSTSTSPVQPTTP---SCTAERWACCGGNGSGCTTCVAGSTCTKINDMYHQCL 297

RESULT 9
US-10-007-521-22

? Sequence 22, Application US/10007521
? Publication No. US20030054539A1
? GENERAL INFORMATION:
? APPLICANT: Schulein, Martin
? Andersen, Lene N.
? Lassen, Soren F.
? Kauppinen, Markus S.
? Lange, Lene
? Nielsen, Ruby I.
? Ihara, Michiko
? Takagi, Shinobu
? TITLE OF INVENTION: No. US20030054539A1el Endoglucanases
? NUMBER OF SEQUENCES: 109
? CORRESPONDENCE ADDRESS:
? ADDRESS: No. US20030054539A1o No. US20030054539A1disk of No. US20030054539A1
? STREET: 405 Lexington Avenue, 64th Floor
? CITY: New York
? STATE: New York
? COUNTRY: United States of America
? ZIP: 10174-6401
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/10/007,521
? FILING DATE: 10-Dec-2001
? CLASSIFICATION: <Unknown>
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/651,136
? FILING DATE: 21-MAY-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Lambiris, Elias J.
? REGISTRATION NUMBER: 33,728
? REFERENCE/DOCKET NUMBER: 4366.200-US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212-867-0123
? TELEFAX: 212-878-9655
? INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 310 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 22
US-10-007-521-22

Query Match 58.0%; Score 962; DB 14; Length 310;
Best Local Similarity 55.9%; Pred. No. 8.7e-75;
Matches 175; Conservative 46; Mismatches 72; Indels 20; Gaps 5;

QY 1 MESTPVLRTTAAALPLVSAASGSGSTRYMDCKKPCSCAMPGRKAAVSQPVYACDANFOR 60
DB 1 MRSAAVLIGLVAG---VAAQSSGHTTRTYMDCKKPCSCAMPGRKAAVSQPVYACDANFOR 56
QY 61 LSPFNQSGC-NGSAYSCADQTPMAVNDNLAYGPAATSIAGSESSGCCACTALFTSG 119
DB 57 LAS-TARSCDMSNVAATCNDQNPMAVNDNLAYGPAATSIAGSESSGCCACTALFTSG 115
QY 120 PVAGKTMVQSTSTGDLGSDNPDIMPGGAGVGFNGCSQFG-GLPQAQVGISSRDQC 178
DB 116 PVAGKTMVQSTSTGDLGSDNPDIMPGGAGVGFNGCSQFG-GLPQAQVGISSRDQC 175
QY 179 DSEPAFLKPGCOMRFPMFONADNPTFTFOQVQCPAETVARSGCKRNDSSFPVTPPSG 235
DB 176 SQLPKALQGCNKRITDMFNADNPDVSMRRYQCPALITDPTGRSDDNVFPVQGPFP 235
QY 236 -----SGANGGTGTPTSTAPSGSGTSPGSSGCTSQKMAQCGGIGTTCVSG 285
DB 236 ATTIRSTTTTASSSSSSSSSTTASGPVFTGGSGPTSPVWQCGGCGMSGPTCAVAG 295
QY 286 TTCOKLNDYSSQC 298
DB 236 STCSVNPWYSQC 308

RESULT 10

US-10-007-521-8
Sequence 8, Application US/10007521
Publication No. US20030054539A1
GENERAL INFORMATION:
APPLICANT: Schulein, Martin

Lassen, Soren F.
Kauppinen, Markus S.
Lange, Lene
Nielsen, Rudy I.
Ihara, Michiko
Takagi, Shinobu

TITLE OF INVENTION: No. US20030054539A1 Endoglucanases
NUMBER OF SEQUENCES: 109
CORRESPONDENCE ADDRESSES:
ADDRESSER: No. US20030054539A1 No. US20030054539A1disk of No. US20030054539A1
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/007,521

CLASSIFICATION: <Unknown>

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US/08/651,136

FILING DATE: 21-MAY-1996

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33, 728

REFERENCE/DOCKET NUMBER: 4366, 200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 295 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-007-521-8

Query Match 57.3%; Score 951; DB 14; Length 295;
Best Local Similarity 58.5%; Pred. No. 6.9e-72;
Matches 166; Conservative 37; Mismatches 78; Indels 10; Gaps 5;

QY 1 MESTPVLRTTAAALPLVSAASGSGSTRYMDCKKPCSCAMPGRKAAVSQPVYACDANFOR 60
DB 1 MRSAAVLIGLVAG---VAAQSSGHTTRTYMDCKKPCSCAMPGRKAAVSQPVYACDANFOR 56
QY 61 LSPFNQSGC-NGSAYSCADQTPMAVNDNLAYGPAATSIAGSESSGCCACTALFTSG 119
DB 57 LAS-TARSCDMSNVAATCNDQNPMAVNDNLAYGPAATSIAGSESSGCCACTALFTSG 115
QY 120 PVAGKTMVQSTSTGDLGSDNPDIMPGGAGVGFNGCSQFG-GLPQAQVGISSRDQC 178
DB 116 PVAGKTMVQSTSTGDLGSDNPDIMPGGAGVGFNGCSQFG-GLPQAQVGISSRDQC 175
QY 179 DSEPAFLKPGCOMRFPMFONADNPTFTFOQVQCPAETVARSGCKRNDSSFPVTPPSG 238
DB 176 AELPSVALDGHRYMFMNDNPNVMMRRVRCPALITNRSGCYRNDNVPYFER---G 232
QY 239 NGGTGTPTSTAPSGSGTSPGSSGCTSQKMAQCGGIGTTCVSGTTCOKLNDYSSQC 298
DB 233 TGPPTPTTPTTSSPQPTKGGGCTSPHWQCGGCGMSGPTACAGGSTCNLNPWYSQC 292
QY 299 L 299
DB 293 I 293

RESULT 11

US-09-735-787-4

Sequence 4, Application US/09735787

Patent No. US20010036910A1

GENERAL INFORMATION:

APPLICANT: Rasmussen, Grethe
Mikkelsen, Jan Moller
Schulein, Martin
Packer, Shankar A.

TITLE OF INVENTION: A Cellulase Preparation Comprising an

Endoglucanase Enzyme

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESSES:
ADDRESSER: No. US20010036910A1 No. US20010036910A1disk of No. US20010036910A1

STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/735,787

CLASSIFICATION: <Unknown>

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US/09/735,787

FILING DATE: 13-Dec-2000

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.

APPLICATION NUMBER: 09/189,028
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3469,214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 376 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-735-787-4

Query Match 56.6%; Score 939.5; DB 9; Length 376;
Best Local Similarity 49.5%; Pred. No. 8,5e-71;
Matches 183; Conservative 35; Mismatches 73; Indels 79; Gaps 6;

QY 9 TTTAALPLVSAASGSGSTRVYDCCCKPCSCAPKAASQPVYACDANFQRLSDPFWQS 68
DB 5 TTTAALPLVSAASGSGSTRVYDCCCKPCSCAPKAASQPVYACDANFQRLSDPFWQS 64
QY 69 GC-NGSAVSCADQTPMAVNDLAVGPAATSIAGSESSWCACALTTTSPVAGKTMV 127
DB 65 GEGGGSAYACTNYSPPAVNDELAVGPAATSIAGSESSWCACALTTTSPVAGKTMV 124
QY 128 VQSTGTGDLGNSQFDLAMPGGVGIFFGCTSEFGKALGAGVGISSRSCDYPFLK 186
DB 125 VQSTGTGDLGNSQFDLAMPGGVGIFFGCTSEFGKALGAGVGISSRSCDYPFLK 184
QY 187 PCGWRPFQNDNPTFTFOVOCPEIIVASGCKRNDSSFPVF----- 232
DB 185 DGHWRPFQNDNPTFTFOVOCPEIIVASGCKRNDSSFPVF----- 232
QY 233 -----TPPS-----GNGGTGTPSTAPSGQTSF 257
DB 245 AKTTTAAAAAQPKTQKDSAPVQKSTKPAQAQPEPTKPADKQTDKPVATKPAATKPVQ 304
QY 258 GGGSGCTSQK-----WAQCGGIGF-----SGCTCVSGTTCC 289
DB 305 PUNKKPTQKVRGTRKSCCAKTDATAKASVPAYVCGGSSKSAVPNGMLACATGSKCV 364
QY 290 KUNDYYSQCL 299
DB 365 KUNYYSQCV 374

RESULT 12
US-10-138-870-4
; Sequence 4, Application US/10138870
; Publication No. US20030119167A1
GENERAL INFORMATION:
APPLICANT: Rasmussen, Gretche
Mikkelsen, Jan Moller
Schulein, Martin
Patkar, Shankant A.
Hagen, Fred
TITLE OF INVENTION: A Cellulase Preparation Comprising an
Endoglucanase Enzyme
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. US20030119167A1; No. US20030119167A1disk of No. US20030119167A1
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIC TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10138,870
FILING DATE: 03-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/735,787
FILING DATE: 13-Dec-2000
APPLICATION NUMBER: 09/189,028
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3469,214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 376 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-138-870-4

Query Match 56.6%; Score 939.5; DB 14; Length 376;
Best Local Similarity 49.5%; Pred. No. 8,5e-71;
Matches 183; Conservative 35; Mismatches 73; Indels 79; Gaps 6;

QY 9 TTTAALPLVSAASGSGSTRVYDCCCKPCSCAPKAASQPVYACDANFQRLSDPFWQS 68
DB 5 TTTAALPLVSAASGSGSTRVYDCCCKPCSCAPKAASQPVYACDANFQRLSDPFWQS 64
QY 69 GC-NGSAVSCADQTPMAVNDLAVGPAATSIAGSESSWCACALTTTSPVAGKTMV 127
DB 65 GEGGGSAYACTNYSPPAVNDELAVGPAATSIAGSESSWCACALTTTSPVAGKTMV 124
QY 128 VQSTGTGDLGNSQFDLAMPGGVGIFFGCTSEFGKALGAGVGISSRSCDYPFLK 186
DB 125 VQSTGTGDLGNSQFDLAMPGGVGIFFGCTSEFGKALGAGVGISSRSCDYPFLK 184
QY 187 PCGWRPFQNDNPTFTFOVOCPEIIVASGCKRNDSSFPVF----- 232
DB 185 DGHWRPFQNDNPTFTFOVOCPEIIVASGCKRNDSSFPVF----- 232
QY 233 -----TPPS-----GNGGTGTPSTAPSGQTSF 257
DB 245 AKTTTAAAAAQPKTQKDSAPVQKSTKPAQAQPEPTKPADKQTDKPVATKPAATKPVQ 304
QY 258 GGGSGCTSQK-----WAQCGGIGF-----SGCTCVSGTTCC 289
DB 305 PUNKKPTQKVRGTRKSCCAKTDATAKASVPAYVCGGSSKSAVPNGMLACATGSKCV 364
QY 290 KUNDYYSQCL 299
DB 365 KUNYYSQCV 374

RESULT 13
US-09-261-329-4
; Sequence 4, Application US/09261329
; Publication No. US20030092097A1
GENERAL INFORMATION:
APPLICANT: Andersen, Kim
Schulein, Martin
Patkar, Shankant A.
Hagen, Fred
TITLE OF INVENTION: A Cellulase Preparation Comprising an
Endoglucanase Enzyme
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. US20030092097A1; No. US20030092097A1disk of No. US20030092097A1
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIC TYPE: Floppy disk

CURRENT APPLICATION NUMBER: US/09/261,329
CURRENT FILING DATE: 1999-03-03
EARLIER APPLICATION NUMBER: 1013/96
EARLIER FILING DATE: 1996-09-17
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 4
LENGTH: 201
TYPE: PRC
ORGANISM: Cellulase variants
US-09-261-329-4

Query Match 56.5%; Score 938; DB 10; Length 201;
Best Local Similarity 79.0%; Pred. No. 5,6e-71;
Matches 158; Conservative 26; Mismatches 16; Indels 0; Gaps 0;

QY 24 GSGSTRYWDCKPCSCAMPKRAVSPYVACDANFOLSDPENVQSGNGGSAVSCADQTP 83
1 GSGSTRYWDCKPCSCAMPKRAVSPYVACDANFOLSDPENVQSGNGGSAVSCADQTP 60
QY 84 WANDLAFPAATSIAGSSSSWCCACALFTTSPVAKTMVOSTSTGDLGSHPD 143
61 WANDLAFPAATSIAGSSSSWCCACALFTTSPVAKTMVOSTSTGDLGSHPD 120
QY 144 IAMPGGVGFNCGSSQFGLPGAQYGGISSRDQDSFPAPLKPCCQRRPFWQNDPFT 203
121 IAMPGGVGFNCGSSQFGLPGAQYGGISSRDQDSFPAPLKPCCQRRPFWQNDPFT 180
QY 204 PTFQOVQCPAEIVARSQCKR 223
181 PTFQOVQCPAEIVARSQCKR 200
DB

RESULT 14
US-10-007-521-24
Sequence 24, Application US/10007521
Publication No. US20030054539A1
GENERAL INFORMATION:
APPLICANT: Schuelein, Martin
Andersen, Lene N.
Lassen, Soren F.
Kauppinen, Markus S.
Lange, Lene
Nielsen, Ruby I.
Ihara, Michiko
Takagi, Shinobu

TITLE OF INVENTION: No. US20030054539A1 Endoglycanases
NUMBER OF SEQUENCES: 109
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. US20030054539A10 No. US20030054539A1disk of No. US20030054539A1
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/007,521
FILING DATE: 10-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/651,136
FILING DATE: 21-MAY-1996

ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4366.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 294 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-10-007-521-24

Query Match 55.7%; Score 923.5; DB 14; Length 294;
Best Local Similarity 57.9%; Pred. No. 1.4e-69;
Matches 169; Conservative 43; Mismatches 69; Indels 11; Gaps 6;

QY 13 ALPLVVASASGSGSTRYWDCKPCSCAMPKRAVSPYVACDANFOLSDPENVQSGC 70
9 SALPLPFAATSIAGSSSSWCCACALFTTSPVAKTMVOSTSTGDLGSHPD 67
QY 71 NGGSAYSCADQTPWANDLAFPAATSIAGSSSSWCCACALFTTSPVAKTMVOSTSTGDLGSHPD 130
68 DGSAYVCSNQGPWANDLAFPAATSIAGSSSSWCCACALFTTSPVAKTMVOSTSTGDLGSHPD 127
QY 131 TSTGDLGSHPD IAMPGGVGFNCGSSQFGLPGAQYGGISSRDQDSFPAPLKPCC 183
128 TSTGDLGSHPD IAMPGGVGFNCGSSQFGLPGAQYGGISSRDQDSFPAPLKPCC 187
QY 190 QRRPFWQNDPFT PTFQOVQCPAEIVARSQCKRNDSD--SFVFTPEPSGNGGTGTPFTS 247
188 QRRPFWQNDPFT PTFQOVQCPAEIVARSQCKRNDSD--SFVFTPEPSGNGGTGTPFTS 245
QY 248 TAPSGGQTSFGSGSGCTSGXMAQCGGIGFSGCTTCVSGTTCQKINDYSGCT 299
246 SSPVQPTT--SGCTARMAQCGGNGGCTTCVASTCTCXINDYHQC 294
DB

RESULT 15
US-10-007-521-10
Sequence 10, Application US/10007521
Publication No. US20030054539A1
GENERAL INFORMATION:
APPLICANT: Schuelein, Martin
Andersen, Lene N.
Lassen, Soren F.
Kauppinen, Markus S.
Lange, Lene
Nielsen, Ruby I.
Ihara, Michiko
Takagi, Shinobu

TITLE OF INVENTION: No. US20030054539A1 Endoglycanases
NUMBER OF SEQUENCES: 109
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. US20030054539A10 No. US20030054539A1disk of No. US20030054539A1
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/007,521
FILING DATE: 10-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/651,136
FILING DATE: 21-MAY-1996

ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4366.200-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 349 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULAR TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-10-007-521-10

Query Match

54.9%; Score 910.5; DR 14; Length 349;

Best Local Similarity 50.1%; Pred. No. 2,1e-68;

Matches 168; Conservative 41; Mismatches 77; Indels 49; Gaps 6;

```
QY 11 LAALFLVASAASGSGSTRWDCCKPSCAMPKAAVSQPVYACDANFORLSDPNVQSGC 70
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10 LVGAVPSSVMAASGKHTRWDCCKTSCKMEGKASVSEPVITCKQDNPIVDANARSGC 69
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 71 NGSAYSCADQTPMAVNDNLAYFPAISIAQSSSSMCCACVALFTSGPVAGKTWVQS 130
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70 DGGAFACITNSPMAVSEDLAYFAFATLSGTEGSMCCACVALFTSGPVAGKTWVQS 129
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 131 TSTGELISNOFDIAMPGGVGITENGCSORFG-LPGAQYGISRDCDSPPAPLKPGC 189
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 130 TITGDLISNNHFDLIPGGGLIPIGSCNPGQLPGRYGVSSRSQCDMPKELIKDQC 189
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 190 QWEFDMPQNALNPFTTFCQVQCPAELIVARSGCKRNDSSFPVETTPSGGNGGTGP--- 245
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 190 QWRFDMPQNSINPDIETFEVQCPKELIVSGCYRDDSSFPVE--QSGSGDVNPPPKPT 247
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 246 -----TSTAPG-----SGQISPGG-----GSQCT 264
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 248 TTTSKPKTISAPSTISNPAPQPGWTDPAETTTTKLPALPATISSPAVSPSSAR 307
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 265 SQMAQCGG-IGFGCTTCVSGTTCQKLDVYSQC 298
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 308 VPIWGQCDSEASWDAEKKCAKGIKCVVNDVYSQC 342
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: July 7, 2004, 08:47:50
Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using SW model

Run on: July 7, 2004, 08:48:30 ; Search time 21 Seconds
(without alignments)
1369.584 Million cell updates/sec

Title: US-10-007-521-12

Perfect score: 1659

Sequence: 1 MRSIPVLRITAAALPLVAS.....TTVCSTGCTCKKNDYISQCL 299

Scoring table: BLOSUM62

Gapop 10.0, Capext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	740.5	44.6	229	2 UC7308	cellulase (EC 3.2.1.4)
2	507	30.6	511	2 S10527	endoglucanase B pr
3	462	27.8	393	2 S59499	cellulase egII - s
4	221	13.3	242	2 S60143	cellulase (EC 3.2.1.4)
5	219.5	13.2	513	1 EUNOI	cellulose 1,4-beta
6	219.5	13.2	513	2 S45380	cellulose 1,4-beta
7	215.5	13.0	513	2 S11439	cellulose 1,4-beta
8	214	12.9	459	2 A25928	cellulase (EC 3.2.1.4)
9	203	12.2	463	2 A48375	cellulase (EC 3.2.1.4)
10	190	11.5	516	2 S33164	cellulase (EC 3.2.1.4)
11	189	11.4	464	2 UC7143	cellulase I -
12	188.5	11.4	302	2 S71334	acetyl xylan ester
13	185	11.1	525	1 S38734	cellulase 1,4-beta
14	184	11.1	516	2 UC5083	cellulase 1,4-beta
15	183	11.0	537	1 UC0150	cellulase 1,4-beta
16	181	10.9	516	2 S42093	cellulase 1,4-beta
17	172	10.3	320	2 UC1311	cell protein precu
18	170.5	10.3	510	2 S41943	cellulase 1,4-beta
19	170.5	10.3	511	2 S44716	cellulase 1,4-beta
20	161	9.7	504	1 B48939	cellulase 1,4-beta
21	156	9.4	856	2 T00349	Avicelase III - As
22	150.5	9.1	860	2 S43846	xylanase 3 - rumen
23	143.5	8.6	388	1 UC5461	cellulase (EC 3.2.1.4)
24	143.5	8.5	388	1 S43920	cellulase (EC 3.2.1.4)
25	141.5	8.4	530	2 UC7979	cellulobiohydrolase
26	139.5	8.3	540	2 S41942	cellulase 1,4-beta
27	137	8.3	438	1 S70602	cellulose 1,4-beta
28	137	8.3	839	2 F75518	hypothetical prote
29	129	7.8	418	1 S28372	cellulase (EC 3.2.1.4)

30	127	7.7	429	1 UC5661	endo-1,4-beta-xyla
31	125.5	7.6	316	1 A38743	loricin - human
32	125.5	7.6	410	1 S68153	cellulase (EC 3.2.1.4)
33	124	7.5	471	1 A26160	cellulose 1,4-beta
34	124	7.5	471	1 A38979	cellulose 1,4-beta
35	122	7.4	182	2 A36686	ultra-high-sulfur
36	120.5	7.3	303	2 H70716	hypothetical prote
37	120	7.2	5376	2 T42215	vestigial protein
38	118.5	7.1	453	2 A41640	triphosphin - human
39	118.5	7.1	749	2 I38488	cellobiohydrolase
40	117	7.1	459	2 UC7931	hypothetical glyci
41	116.5	7.0	584	2 G70804	keratin KAP5.4 - s
42	115.5	7.0	191	2 T46412	related to spore c
43	114	6.9	461	2 T51044	hypothetical prote
44	114	6.9	1804	2 H96597	hypothetical glyci
45	113.5	6.8	1381	2 E73806	hypothetical glyci

ALIGNMENTS

RESULT 1

UC7308 cellulase (EC 3.2.1.4) - Scopulariopsis brevicaulis

M:Alternate names: endoglucanase I

C:Species: Scopulariopsis brevicaulis

C>Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000

C:Accession: UC7308; EC7087

R:Nakatsuji, F.; Kawaguchi, T.; Takada, G.; Sumitani, J.; Moriyama, Y.; Arai, M.

BioSci. Biotechnol. Biochem. 64, 1338-1346, 2000

A:Title: Cloning and sequencing of an endoglucanase gene from Scopulariopsis brevicaulis

A:Reference number: UC7308

A:Accession: UC7308

A:Molecule type: DNA

A:Residues: 1-229 <NAK>

A:Experimental source: strain TOF-12:2

A:Accession: PC7087

A:Molecule type: Protein

A:Residues: 21-37,149-164 <NA2>

A:Gene: egi

A:Initiators: 147/3

C:Keywords: glycosidase; hydrolase

Query Match 44.6%; Score 740.5; DB 2; Length 229;

Best Local Similarity 59.8%; Pred. No. 2.6e-45;

Matches 134; Conservative 32; Mismatches 49; Indels 9; Gaps 5;

QY	10	TLAAALPLV---ASMAAGSGQSTRYWDCKPSCAMPGRRAVSQ-PYVACDANFOLSDFN 65
DB	6	TLALTLFLVLPASQSGTGTRTYMDCCKPSCMPDRAPLSQGPMTCDINDNPLDDGG 65
QY	66	V-GSGCC-GGASACADOTPMANVMDIAYGPAATISJAGSSSSWCACACALFTTSGPAG 123
DB	66	LTESGCPGGGAGACSSSPMAVDDDELAYGMAAVNIQQTSSDWCACALFTTGAIVSG 125
QY	124	KTMVQSTSTGDIAGNODLAMPGGVGLPFGCCSSQFGLP---GAQYGISSDQCD 180
DB	124	KMTVQKTMVQSTSTGDIAGNODLAMPGGVGLPFGCCSSQFGLP---GAQYGISSDQCD 180
QY	181	FPAPLKPCQMRPDMFONADNPFTFGQVQCPAEIVARSGCKRN 224
DB	186	FPBALKAGCEMRDWMGTGNDPVSFREVCPAEIVQSGQCRS 229

RESULT 2

S10527 endoglucanase B precursor - Pseudomonas fluorescens

C:Species: Pseudomonas fluorescens

C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999

C:Accession: S10527

R:Gilbert, H.J.; Hall, J.; Hazlewood, G.P.; Ferreira, L.M.A.

Mol. Microbiol. 4, 759-767, 1990

A>Title: The N-terminal region of an endoglucanase from *Pseudomonas fluorescens* subsp. A
A/Reference number: S10527; MUID:90355836; PMID:2117653
A/Accession: S10527
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-511 <GIL>
A/Cross-references: EMBL:X52615; NID:945497; PIDD:CA16844.1; PID:945498
C/Superfamily: bacterial cellulose-binding domain homology; glycosylase GWC domain hom
F:31-128/Domain: bacterial cellulose-binding domain homology <BCB>
F:180-217/Domain: glycosylase GWC domain homology <GWC>
F:32-127/Disulfide bonds: #status predicted

Query Match 30.6%; Score 507; DB 2; Length 511;
Best Local Similarity 39.6%; Pred. No. 1.5e-28;
Matches 107; Conservative 38; Mismatches 73; Indels 52; Gaps 7;

QY 2 RSTPVLRTTAAALPLVNSASG-----SGQSTRYMDCKPSCAPGKA-AVSPQVY 52
DB 227 RSVASSSSSLSATSSASVSPETDGGNGVATRYMDCKPCHGMSANVPLVSPQ 296
QY 53 ACDANFRLSDPNVSGGNSAYSCADQTPAVVNDLNGFPAITSIAGSSSSWCCATCY 112
DB 297 SCSSANWRLSDVSSGCDGGGCMCMIPNAVSPITAVGAAV-----SSGVCGRCY 351
QY 113 ALPFT-----SGPVAGTMYVOSTGTGDLGNSQFDIAMPGGGVGIFNGCSSQF 161
DB 352 QLOPSTSSYNAPDGDGSAALAGKMTIVQATNIGDVSGGQFDIVGGGVGAPACSAQW 411
QY 162 G--GLPGAQYGGISSR-----DQCDSF-----PAPKPGCQWRD 194
DB 412 GVSNAELGAQYGGFLAACKQQLAGYNASTLQYKSCVLRNCDSPVSGRLTQLQGCGTFAF 471
QY 195 WFOANDPPTTTCQVQCPAETIVASGCKEN 224
DB 472 WFEADNPISLTKKEVCPAELTTRSGMRS 501

RESULT 3
S59499
cellulase egli - smut fungus (Ustilago maydis)
C/Species: Ustilago maydis (corn smut)
C/Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 28-May-1999
C/Accession: S59499
R/Schaecker, F.; Wanner, G.; Kahmann, R.
Biol. Chem. Hoppe-Seyler 376, 617-625, 1995
A/Title: Filament-specific expression of a cellulase gene in the dimorphic fungus *Ustilago*
A/Reference number: S59499; MUID:96145728; PMID:8590631
A/Accession: S59499
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-393 <SCH>
A/Cross-references: GB:S81598; NID:g1478378; PIDD:AA836147.1; PID:g1478379

Query Match 27.8%; Score 462; DB 2; Length 393;
Best Local Similarity 30.9%; Pred. No. 1.8e-25;
Matches 115; Conservative 41; Mismatches 100; Indels 116; Gaps 12;

QY 11 LAALPLVNSASGSGQSTRYMDCKPSCAPGKAASQPTACDANPRL-----SPRY 66
DB 12 LSLSLVHLDGVRAGMATRYMDCLASMSGKAPVAPYADCKADGVTLLDSKDPG 71
QY 67 QSGNGGSAVSCADQTPW--AVNDNLAVGPAITSIAGSSSSWCCAYALFTSPQVAK 124
DB 72 QSGNGGSAVSCADQTPW--AVNDNLAVGPAITSIAGSSSSWCCAYALFTSPQVAK 124
QY 125 TW-----VVGSTGTGDLGNSQFDIAMPGGGVGIF--NGCSSQF--GLPGAQYGGISSR 175
DB 129 AMKNKILFQVTVVGGDVQSQNPFQIPGGIGAPKCPAQMGVVASIMDQYGVGYSA 188
QY 176 DQCSFPAPLKGCCQWRP--DMFOANDPPT--TFOVQCCPAETIVASGCKENSDSPVE 232
DB 189 TEGSKLPKPLQEGCKMRFSEW---GDNPVKSGPKAVKCPKSLIDSSGCKRQKDDNTTSPY 245

QY 233 T----- 233
DB 246 SKKIDSANTAAPQYKRBRSVCLAGKKKSAAGVSDSGASGALASGAAABSGSG 305
QY 234 -----PPSGGN-----GGTGTPTSTAPSGQTSPPQGG 260
DB 306 QPEGVCGPSGANDQSSNGDATTGASGSGSDSGSTANGSGAPTSGSASVAPPSGG 365
QY 261 SGGTSQKMAQCG 272
DB 366 SNPGAQGGQGG 377

RESULT 4
S60143
cellulase (EC 3.2.1.4) V precursor - fungus (Trichoderma reesei) (strain QM9414)
N/Alternate names: endo-1,4-beta-glucanase V
C/Species: Trichoderma reesei
A/Variety: strain QM9414
C/Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 17-Mar-1999
C/Accession: S60143; S49043
R/Salohimo, A.; Henriksen, B.; Hoffren, A.; Telemann, O.; Penttila, M.
submitted to the EMBL Data Library, May 1994
A/Description: A novel small endoglucanase gene egli5 from *Trichoderma reesei* isolated
A/Reference number: S60143
A/Accession: S60143
A/Molecule type: DNA
A/Residues: 1-242 <SNL>
A/Cross-references: EMBL:Z33381; NID:9485863; PID:9485864
R/Salohimo, A.; Henriksen, B.; Hoffren, A.M.; Telemann, O.; Penttila, M.
Mol. Microbiol. 13, 219-228, 1994
A/Title: A novel, small endoglucanase gene, egli5, from *Trichoderma reesei* isolated by
A/Reference number: S49043; MUID:95075308; PMID:7984103
A/Accession: S49043
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 206-241 <SMW>
A/Cross-references: EMBL:Z33381
C/Genetics:
A/Gene: egli5
A/Insertions: 23/3; 45/3
C/Function:
A/Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A/pathway: cellulose degradation
C/Superfamily: fungal cellulose-binding domain homology
C/Keywords: glycosylase, hydrolase, polysaccharide degradation
F:1-17/Domain: signal sequence #status predicted <Sig>
F:18-242/Product: cellulase #status predicted <Mat>
F:210-241/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 13.3%; Score 221; DB 2; Length 242;
Best Local Similarity 24.7%; Pred. No. 1.1e-08;
Matches 78; Conservative 39; Mismatches 101; Indels 98; Gaps 13;

QY 7 LRTTLAALPLVNSASGSGQSTRYMD-----CKPSCAPGKAASQPTACDANPQ 59
DB 1 MKATLVLSLIVGAVSAVYATTRRYDQEGACGSSSGAPWGLGIMGVYTR--AGSQ 59
QY 60 RLSDPNVSGGNSAYSCADQTPAVVNDLNGFPAITSIAGSSSSWCCAYALFT-- 117
DB 60 ALFDTAGASWCGAG-----CGKCYQLSTQ 85
QY 118 -----SGPVAGTMYVOSTGTGDLGNSQFDIAMPGGGVGIFNGCSSQFGLPGAQY 170
DB 86 APSSSGTGGAGAGSITVWTVLCPNNGNMQWCPVY--GGTN-----QY- 127
QY 171 GISSRDQDSEFPADLPKGCQWRP--WFOAN--ADNPTFTFQVQCPAETIVASG--CKR 223
DB 128 -----GYSYHFDIQAQNEIFGDNVAVVFBPIACPGQASDWGTCLCVG 170
QY 224 NDSSFPVFTPPSGANGGTPTSTAPSGQTSPPGSSGCTSQKMAQCGGIGSGCTTCV 283
DB 171 QQRTD-P--TVVANDTGSTIPRESSPPATSSPSGGGQQT--LYGCGGAGMTGFTTCQ 225

QY 284 SGTTCQKNDYSSQCL 299
 DB 226 ADGTCKVQNMWYSQCL 241

RESULT 5

cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) I precursor - fungus (Trichoderma reesei)
 N.Alternate names: exo-cellulohydrolase I
 C.Species: Trichoderma reesei
 C.Date: 03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change 16-Jul-1999

C.Accession: A00902
 R.Shemaker, S.; Schweickart, V.; Ladner, M.; Gelfand, D.; Kwok, S.; Myambo, K.; Innis, B./Technology 1, 691-696, 1983
 A.Title: Molecular cloning of exo-cellulohydrolase I derived from Trichoderma reesei
 A.Reference number: A00902
 A.Accession: A00902
 A.Molecule type: DNA
 A.Residues: 1-513 <SHO>
 A.Experimental source: strain 127
 A.Comment: This is the most abundantly produced cellulase in this filamentous fungus; in C.Gene: CHH1
 A.Introns: 154/2; 386/3
 C.Function: A.Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in cellulose to D-glucose
 C.Keywords: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain EC03
 F.1-17/Domain: signal sequence #status predicted <SIG>
 F.18-513/Product: cellulose 1,4-beta-cellulobiosidase I #status predicted <MAT>
 F.482-513/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 13.2%; Score 219.5; DB 1; Length 513;
 Best Local Similarity 26.9%; Pred. No. 2.7e-08;
 Matches 91; Conservative 36; Mismatches 122; Indels 89; Gaps 22;

QY 16 PLYASAGSGSTRYWDCCCKPSGAPGAASQPV--YACDANFQRLDFHWQSGCNG 73
 DB 211 PSSNNANTGIG--GHGSCSEMDIWEAN-SISEALTPHETTVGGEICE--GDGC--G 261
 QY 74 SAY-----C---ADQTPMAVNDLAVGPAATSIAGSSSSSCACVALLFTSGPVA 122
 DB 262 GTTSDNRVGGTCDPDCGMNRYRLGNTSFTG-----PGSS-----FTLDITK----- 303
 QY 123 GKTWVQSTSTGDLG---SNQFDIAMPGGVGIENG-----CS---SQFGGLPQAQ 168
 DB 304 -KLTVTQFETSGAINRYVQNGVTGQPMALISYSGNELNDYCTAEAEAEFGSSPSD 362
 QY 169 YGSISSRDQDSFPAPLKPCCQKRFPMFQN---ADNPTFTQVQCPAEIYASGCKRND 225
 DB 363 KGGITQFKKATSGGAVLWMSL-WD-DYANMLMDSTYPTNERTSTPGAV--RSGCSTSS 418
 QY 226 -----DSSFP-----VFTP-----PSGDN---GGTGTPTSPAGSGGTSPPGGS 261
 DB 419 GVPAYQVESQSPNAKATPSNKKFPIGISTGNSSGKPPGKGRGTTTRRPA---TTTSSSP 475
 QY 262 GCTSQKMAQCGGIGSGGCTTCVSGTTCCMLNDYSSQCL 299
 DB 476 GPTQSHYQCGGIGSGGTPVCASGTTQVLMNPYSQCL 513

RESULT 6

cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) - fungus (Trichoderma koningi)
 C.Species: Trichoderma koningi
 C.Date: 20-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
 C.Accession: S45380
 R.Wey, T.T.; Hseu, T.B.; Huang, L.
 Curr. Microbiol. 28, 31-39, 1994
 A>Title: Molecular cloning and sequence analysis of the cellobiohydrolase I gene from Trichoderma koningi
 A.Reference number: S45380; MUID:94100788; PMID:7764306
 A.Accession: S45380

A.Molecule type: DNA
 A.Residues: 1-513 <MBY>
 A.Cross-References: EMBL:X69976; NID:G457422; PIDD:CAA9596.1; PID:G457423
 C.Gene: CHH1
 A.Introns: 154/2; 386/3
 C.Function: A.Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in cellulose to D-glucose
 C.Keywords: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain EC03
 F.1-17/Domain: signal sequence #status predicted <SIG>
 F.18-513/Product: cellulose 1,4-beta-cellulobiosidase I #status predicted <MAT>
 F.482-513/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 13.2%; Score 219.5; DB 2; Length 513;
 Best Local Similarity 26.9%; Pred. No. 2.7e-08;
 Matches 91; Conservative 36; Mismatches 122; Indels 89; Gaps 22;

QY 16 PLYASAGSGSTRYWDCCCKPSGAPGAASQPV--YACDANFQRLDFHWQSGCNG 73
 DB 211 PSSNNANTGIG--GHGSCSEMDIWEAN-SISEALTPHETTVGGEICE--GDGC--G 261
 QY 74 SAY-----C---ADQTPMAVNDLAVGPAATSIAGSSSSSCACVALLFTSGPVA 122
 DB 262 GTTSDNRVGGTCDPDCGMNRYRLGNTSFTG-----PGSS-----FTLDITK----- 303
 QY 123 GKTWVQSTSTGDLG---SNQFDIAMPGGVGIENG-----CS---SQFGGLPQAQ 168
 DB 304 -KLTVTQFETSGAINRYVQNGVTGQPMALISYSGNELNDYCTAEAEAEFGSSPSD 362
 QY 169 YGSISSRDQDSFPAPLKPCCQKRFPMFQN---ADNPTFTQVQCPAEIYASGCKRND 225
 DB 363 KGGITQFKKATSGGAVLWMSL-WD-DYANMLMDSTYPTNERTSTPGAV--RSGCSTSS 418
 QY 226 -----DSSFP-----VFTP-----PSGDN---GGTGTPTSPAGSGGTSPPGGS 261
 DB 419 GVPAYQVESQSPNAKATPSNKKFPIGISTGNSSGKPPGKGRGTTTRRPA---TTTSSSP 475
 QY 262 GCTSQKMAQCGGIGSGGCTTCVSGTTCCMLNDYSSQCL 299
 DB 476 GPTQSHYQCGGIGSGGTPVCASGTTQVLMNPYSQCL 513

RESULT 7

cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) - fungus (Trichoderma viride)
 C.Species: Trichoderma viride
 C.Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
 C.Accession: S11439
 R.Ohng, C.; Tsukagoshi, N.; Ueda, S.
 Nucleic Acids Res. 18, 5559, 1990
 A>Title: Nucleotide sequence of the cellobiohydrolase gene from Trichoderma viride.
 A.Reference number: S11439; MUID:91016856; PMID:2216737
 A.Accession: S11439
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-513 <CEB>
 A.Cross-References: EMBL:X53931; NID:G5156; PIDD:CAA37878.1; PID:G299337
 C.Gene: CHH1
 A.Introns: 154/2; 386/3
 C.Function: A.Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in cellulose to D-glucose
 C.Keywords: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain EC03
 F.1-17/Domain: signal sequence #status predicted <SIG>
 F.18-513/Product: cellulose 1,4-beta-cellulobiosidase I #status predicted <MAT>
 F.482-513/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 13.0%; Score 215.5; DB 2; Length 513;
 Best Local Similarity 26.2%; Pred. No. 5.2e-08;
 Matches 86; Conservative 36; Mismatches 137; Indels 69; Gaps 16;

QY 16 PLYASAGSGSTRYWDCCCKPSGAPGAASQPV--YACDANFQRLDFHWQSGCNG 73
 DB 211 PSSNNANTGIG--GHGSCSEMDIWEAN-SISEALTPHETTVGGEICE--GDGC--G 261
 QY 74 SAY-----C---ADQTPMAVNDLAVGPAATSIAGSSSSSCACVALLFTSGPVA 122
 DB 262 GCTSQKMAQCGGIGSGGCTTCVSGTTCCMLNDYSSQCL 299
 QY 129 VQSTSTGDLG---SNQFDIAMPGGVGIENG-----SQFGGLPQAQ 173

Db 308 VTQETSGAINRYVQNGVTEQCPNAELGDVSGNSLDDYCAEABFRCSSSFDKGGTL 367
QY 174 SRDCCSPAPLKECCQMPDPQON---ADNFTTFQVCCPAEI---VARSCCKAND 226
Db 368 QPKATSGZMWLVMSL-WD-DYYANMLMDSTVPTDSTSPGAVSGSSSTSSGVPAQLE 425
QY 227 SSFP-----VFTP-----PSGNGSTGTPTSTAPSGGTSQSGSGCTGQMAQC 271
Db 426 SNSPMKAVVYSNIRKPIGSGTGNPSGNGPAGPPTTPPATSTGSSPGTGTTHGQC 485
QY 272 GGIGFSGCTTGVSGTTCOKLNDYVSQCL 299
Db 486 GGIGTGTTCVAGSSSTCCQYLAFFYSQCL 513

RESULT 8

cellulase (EC 3.2.1.4) A precursor - fungus (Trichoderma reesei)
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase I
C:Species: Trichoderma reesei
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 22-Jun-1995
C:Accession: A25928; A25565
R:Van Arsdale, J.N.; Kwock, S.; Schweickart, V.L.; Ladner, K.B.; Gelfand, D.H.; Innis, M.
Bio/Technology 5, 60-64, 1987
A:Title: Cloning, characterization, and expression in Saccharomyces cerevisiae of endogly-
A:Reference number: A25928
A:Accession: A25928
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-459 <VAN>
A:Experimental source: strain 127
R:Peritila, M.; Lehtovaara, P.; Nevalainen, H.; Bhikhabhai, R.; Knowles, J.
Gene 49, 253-263, 1986
A:Title: Homology between cellulase genes of Trichoderma reesei: complete nucleotide seq-
A:Reference number: A25565; MUID:87106822; PMID:2948877
A:Accession: A25565
A:Molecule type: mRNA
A:Residues: 1-459 <PEN>
A:Cross-references: GB:M15665; NID:gl70546; PIDD:AAA34212.1; PID:gl70547
C:Genetics:
A:Gene: eg11
A:Introns: 257/2; 454/1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce-
A:Pathway: cellulose degradation
C:Superfamily: cellulose 1,4-beta-cellulohydrolase I; fungal cellulose-binding domain hom-
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-459/Product: cellulase A #status predicted <MAT>
F:428-459/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 12.9%; Score 214; DB 2; Length 459;
Best Local Similarity 27.0%; Pred. No. 6e-08;
Matches 108; Conservative 31; Mismatches 101; Indels 160; Gaps 27;

QY 19 ASAASGSGS---TRWDCCKPCAMPKAAVSOPVYACD-----NQQLSPFN 66
Db 101 ASGVTSSSSLTMYQV---WPSSS-GGYSVSPRLYLDDSDGYMLKLNQDEL 154
QY 67 Q-----SCNGSAYSCADQTP---NAVNDNIAYGFA 95
Db 155 DLSALPGENGSLYLSQMDENGANGVNTAGVSGY-CAQCPVQWTR-NGTL----- 207
QY 96 ATSTAGSSESSMCC-----AC-----YALFTS-GPV 121
Db 208 -----NTSHQFPCNEMDILBGNBRANALTPHSCATATCDASGCFNPGSGYKSYGP- 261
QY 122 AGKTMVQST-----STGCDLGS-----NQPD1-AMPGGVGFNPGSSQ 160
Db 262 -GDYVDSKTFITITQFNTDNGSPSGNLVSTRKYGQGVIPAGQGGDT--ISSCS- 317
QY 161 FGGIPLGAQYGGISSRQCDSPAPLKECCQMPDPQONAD-NPFTTFQVCC----- 211

Db 318 -----ASAYGLATMGKA-----LSSGMLVPSIWNDSQYMMWLDGNAQCSSTEGN 366
QY 212 PSEIYASGCKRRDSDSPVFTTPSGNGAGTGTPTSTA-----PSGQT-----SPGG 259
Db 367 PSNIIA-----NNPHTVHFVSNIRMDIDS-PTNSTAPPPPPASSTFTTTRRSSTSS 419
QY 260 GSGCTSGKMAQCGGIGFSGCTTGVSGTTCOKLNDYVSQCL 299
Db 420 SPSCQTQTHMGCGGIGTSGCKTCTSGTTCCQXSNDYVSQCL 459

RESULT 9

cellulase (EC 3.2.1.4) - fungus (Trichoderma longibrachiatum)
A:Accession: A48375
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase I homolog EGL1
C:Species: Trichoderma longibrachiatum
C:Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 22-Jun-1995
C:Accession: A48375; S28521
R:Gonzalez, R.; Ramon, D.; Perez-Gonzalez, J.A.
Appl. Microbiol. Biotechnol. 38, 370-375, 1992
A:Title: Cloning, sequence analysis and yeast expression of the eg11 gene from Trich-
A:Reference number: A48375; MUID:93159747; PMID:1369161
A:Accession: A48375
A:Contents: CECT 2606
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-463 <GON>
A:Cross-references: EMBL:X60652; NID:95181; PIDD:CAA43059.1; PID:95182
A:Note: sequence extracted from NCBI backbone (NCBIN:125157, NCBI:P.125158)
C:Genetics:
A:Introns: 257/2; 458/1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Superfamily: cellulose 1,4-beta-cellulohydrolase I; fungal cellulose-binding domain h
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:432-463/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 12.2%; Score 203; DB 2; Length 463;
Best Local Similarity 28.4%; Pred. No. 3.6e-07;
Matches 86; Conservative 22; Mismatches 101; Indels 94; Gaps 18;

QY 68 SGCNGSAYSCADQTP---NAVNDNL---AYGPAAT-SIAGSSSSMCCACTALPTFSQ 119
Db 184 AGANVGSGY-CAQCPVQWTR-NGTLNTSGGFCNEMDILBGNBRANALTPHSCATAC 241
QY 120 PVAG-----KTMV-----QSTGCDLGS-----NQPD 143
Db 242 DSAAGGFNPGSGYPNYFGGDTVDSKTFITITQFNTDNGSPSGNLVSTRKTRQNGVD 301
QY 144 I--AMPGGVGFNPGSSGCGFGLPGAQYGGISSRDQDSFPAPLKECCQMPDPQONAD 200
Db 302 ITPAKRGDT--ISSCS-----ASAYGLATMGKA-----LSSGMLVPSIWNDSQ 347
QY 201 -----NPTTFQVCCPAIYASGCKRRDSDSPVFTTPSGNGAGTGTPTSTA----- 239
Db 348 YMMWLDGNAQCSSTEGNPSNII-ANNPHTVHFVSNIRMDIDS-PTNSTAPPPPPASSTFTTTRRSSTSS 402
QY 240 ---GCTGPTTADSGQTSPPGSSGCTSGKMAQCGGIGFSGCTTGVSGTTCOKLNDYVS 296
Db 403 PEPASSTFTTTRRSSTSS-SSPSCQTQTHMGCGGIGTSGCKTCTSGTTCCQXNDYVS 460
QY 297 QCL 299
Db 461 QCL 463

RESULT 10

S33164
cellulose 1,4-beta-cellulohydrolase (EC 3.2.1.91) - basidiomycete (Phanerochaete chrysos-
C:Species: Phanerochaete chrysosporium
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1995
C:Accession: S33164

R.Sims, P.F.G.; Soares-Felipe, S.M.; Gent, M.E.; Tempelaars, C.; Wang, Q.; Broda, P.
 Submitted to the EMBL Data Library, April 1993
 A:Description: Differential expression of multiple exo-cellulohydrolase I-like genes in
 A:Reference number: S33164
 A:Accession: S33164
 A:Molecule type: DNA
 A:Residues: 1-516 <S1>
 A:Cross-references: EMBL:Z22528; NID:g296026; PDB:CA8023.1; PDB:g296027
 C:Superfamily: cellulose 1,4-beta-cellulohydrolase I; fungal cellulose-binding domain hom
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:485-516/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 11.5%; Score 190; DB 2; Length 516;
 Best Local Similarity 24.4%; Pred. No. 3.3e-06;
 Matches 88; Conservative 34; Mismatches 101; Indels 138; Gaps 18;

20 SAASGGSTRYMCKPCSCAMPK--AAVSQPYACDANFQLSDPVSQNGG--- 73
 211 SANAGT---NYGTCCTENDMEANNDAAATP-HCCT-----NAQTRCGSSDCR 258
 74 -----SAYSCADQ-----PMAY-----NDNTAYGFATSTAGS 103
 259 FGLGDADGCDPNSFRMGDQTFAGKGLTVDTSKPTVYVQFTINDGTSAG----- 308
 104 ESWCCACVALLTFSGVAKTMVQSTST---GDLGNCFLDAMGGGVGIFNCGSSQ 160
 309 -----TLTEIRLYVQNGKVIQNSVXKIPIDVNSTD-----NFCSSQ 348
 161 ---FGLP-GAQYGGISRRDC-----DSFPAPLRGCGMPFMPQ-----NA 199
 349 KRAFQTYVPAGHGLAQVGEALRTGWTALSTINDVYANML-----WLOSVPYTK 400
 200 LNPFTFQGVOC-----PAEIVASGCKRNDSPFVTPPSGNGGT----- 242
 401 DPETPVARGCATTSQVPAQIAQS-----PVAIVPSNIXFDLNTYTGTVSSSV 454
 243 -----CTPSTARSGQTSFGSSGCTSQKMAQCGIGSGCTTCQKLNYSQ 297
 455 SSSHSSTSSSSSSSTPTPTGTVPQMGCGIGYSTTCASPTCHVLANPYSQ 514

Db 298 C 298
 515 C 515

RESULT 11
 JCT143
 endoglucanase I - fungus (Trichoderma viride)
 C:Species: Trichoderma viride
 C:Date: 04-Mar-2000 #sequence_revision 04-Mar-2000 #text_change 31-Mar-2003
 C:Accession: JCT143
 R:Kwon, I.; Eklino, K.; Gotc, M.; Furukawa, K.
 Biosci. Biotechnol. Biochem. 63, 1714-1720, 1999
 A:Title: Heterologous expression and characterization of endoglucanase I (Egl) from Tric
 A:Reference number: JCT143; MUID:20052947; PMID:10586500
 A:Accession: JCT143
 A:Molecule type: mRNA
 A:Residues: 1-464 <RMC>
 A:Experimental source: HK-75
 C:Comment: This protein is a fusion glycoprotein with catalytic and cellulose binding do
 C:Genetics:
 A:Gene: eg11
 A:Introns: 257/2, 459/2
 C:Superfamily: cellulose 1,4-beta-cellulohydrolase I; fungal cellulose-binding domain hom
 C:Keywords: glycoprotein
 F:433-464/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 11.4%; Score 189; DB 2; Length 464;
 Best Local Similarity 25.4%; Pred. No. 3.5e-06;
 Matches 102; Conservative 32; Mismatches 108; Indels 160; Gaps 25;

21 AAGG---SSQS---TRYMDCKPSCAMPKAAVSQPYACD-----NFORLSQDN 65

Db 100 AASGVQTSALITMNY-----WPSSS-GGYSSVSPRLYLDSGEVYLKINGQLS-FD 153
 66 VQ-----SCNGSAGNYCAQGT---MAVNDL---A 92
 154 VDLALPCEGNSGLYLSMDCTGAGANQVNTAGANGSY-CDAGCVQTR-NGLTNTH 211
 92 YGFAT--SIAGSSSSWCCACVALLTFSGPVAG-----K 124
 212 QGFCEBMDLIGNSRANALTPHSCRTKICDSAGCGRPVSGGYMYRGDTVDSNVF 272
 125 TMVQ-STSYGDLGS-----NORDI--AMPGGVGINSGSSQFGLPCAQYGG 171
 272 TITQNTDGSATNULVSTRKTRQNGVDIPAPRGDT--ISSCS-----ASAVGG 323
 172 ISSDQCDSEFAPLRGCGMPF-WFQND-----NPTFPQVQCP 212
 324 LTTMKR-----LSSGVVLFSTIMDNQYMMMLDSGAPCSSTEGNPT----- 368
 213 AEIVAS-----GCKRNDSPFVTPPSGNGGCTGPTSTAPSGQTSF 257
 369 -NITANPNTHVPSNIRMGDISSTNSGTTP---PPPPPPASSTLSTRBSSTTS 423
 258 GGSGLTSQKMAQCGIGFSQCTTCVSGTTCQKLNYSQCL 299
 424 -SSPCTQTHYGCWNGIGYTCCKTCAAGTTCQYSDYSSQCL 464

Db 258 GGSGLTSQKMAQCGIGFSQCTTCVSGTTCQKLNYSQCL 299
 424 -SSPCTQTHYGCWNGIGYTCCKTCAAGTTCQYSDYSSQCL 464

RESULT 12
 S71334
 acetyl xylan esterase precursor - fungus (Trichoderma reesei)
 C:Species: Trichoderma reesei
 C:Date: 23-Jul-1997 #sequence_revision 01-Aug-1997 #text_change 17-Mar-1999
 C:Accession: S71334
 R:Margolles-Claek, E.; Tenkanen, M.; Soederlund, H.; Penttilae, M.
 Eur. J. Biochem. 237, 553-560, 1996
 A:Title: Acetyl xylan esterase from Trichoderma reesei contains an active-site serine
 A:Reference number: S71334; MUID:96235218; PMID:8647098
 A:Accession: S71334
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-302 <MAR>
 A:Cross-references: EMBL:Z69256; NID:g1431619; PDB:e220701; PDB:g1431620
 C:Genetics:
 A:Gene: axel
 C:Superfamily: fungal cellulose-binding domain homology
 F:1-20/Domain: signal sequence #status predicted <SIG>
 F:21-302/Product: acetyl xylan esterase #status predicted <MAT>
 F:271-302/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 11.4%; Score 188.5; DB 2; Length 302;
 Best Local Similarity 26.6%; Pred. No. 2.5e-06;
 Matches 73; Conservative 17; Mismatches 97; Indels 87; Gaps 8;

92 YGFATSI-----AGSSESSWCCACVALLTFSGPVAGKTMV----- 127
 50 YSSSATVNVNVIQAHRTTSEAIYPAQGGQASCGGISVANSVNGTNAALANNFNHS 109
 128 ---VQSTSTGDLGNSQDPIAPPG---GVIPNCSQFGLPGAQYGGISSRQCSF 181
 110 CPDTQIVLVVYSQAGIIPDNALCGGDPREGITNTAVPLTGAVASAKAALFMGDPNVIH 169
 182 PAPLRGCGMPFMPQNDNPTFPQVQ-----CPAEIVASGCKR----- 223
 170 GLPYNV-----TCTTGSEAPRAPAGVYCSAKIKSYDAADAPCCYGNPD 215
 224 -----NDSSFPVTPPSGNGGCTGPTSTAPSGQTSRQSSGCTS 265
 216 NVHGVGEYGGQNALAFINSQLSGSGSQPPGGFTSTRPTSTRTS---SREP---TO 268
 266 QKMAQCGIGFSQCTTCVSGTTCQKLNYSQCL 299
 269 THMGCGCGGMTPPTQCESGTTTCQVVISQCL 302

C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:506-537/Domain: fungal cellulose-binding domain homology <FCB>

Query Match:

11.04; Score 183; DB 1; Length 537;

Best Local Similarity 23.44; Pred. No. 1.1e-05;

Matches 79; Conservative 30; Mismatches 99; Indels 130; Gaps 14;

```
QY 24 GSGQSRITWDCCKPSCAMPGKAASQPIYACDANFQRLSDRYVSGCNGSAYSACADQTP 83
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 268 GYISIDRYGGTCDEP-----GCFNFPYRMGVTFNY-----GGRTIDTKSP 308

QY 84 WAV-----NDNLAYGPAA-----TSIAGGS-ESSWCCACVALTF 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 303 FIVVDFGLNDGISTGLISEIKRFYVGGKVIQNPQSTIVGSGNSITDSKNA----- 362

QY 117 TSGPVAKMTWVVGSTGTGDLGNOFDI--AMPGGVGIENGQ----- 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 363 -----QKSAFGD--TNEFSKHGGMAGMAGALCGMIVYMSLWDDHASDMLWL 407

QY 158 -----SSQPGGLPGACYGGISRDQDSFPAFLKPGQWRFPQNDNPPTTFQOVQCP 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 408 DSTYPTNATSTTPGAK-----RGTCDISRRP-----NTVESTY-----P 441

QY 213 AEIVARSGCKRNDSSPVPFPTPPSGANGGTGTPSTAPGSGQ-----TSPGGSS 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 442 KAVIYISNKTGPIINS--FTGGTSSSSSTTTTTSKSTSTSSSKITTYTTTTSSSSS 499

QY 262 GCTSGKMAQCGGIGFSGGCTTGVSGITTCQKINDYISQCL 299
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 500 GTGARDMAQCGGNGMTGPTTCVSPYTCTKQNDWYSQCL 537
```

Search completed: July 7, 2004, 08:54:54
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 08:46:50 ; Search time 17 Seconds

(without alignments)
915,822 Million cell updates/sec

Title: US-10-007-521-12

Perfect score: 1659

Sequence: 1 MRSTPVRLTAAALPIVAVS.....TCVSGTTCQKNDYISQCL 293

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	942.5	56.8	1 GUNK_FUSOX	P45699 fusarium ox
2	921.5	55.5	1 GUNS_HUMIN	P43316 humicola in
3	507	30.6	1 GUNB_PSEPL	P18126 pseudomonas
4	462	27.8	1 GUN1_USITMA	P54424 usitlago ma
5	247	14.3	1 PABP_PENRN	Q98e18 penicillium
6	221	13.3	1 GUNS_TRIRE	P43317 trichoderma
7	219.5	13.2	1 GUX1_TRIIV	P00725 trichoderma
8	215.5	13.0	1 GUX1_TRIIV	P19355 trichoderma
9	214	12.9	1 GUN1_TRIRE	P07981 trichoderma
10	203	12.2	1 GUN1_TRIIO	Q12714 trichoderma
11	183	11.2	1 GUX1_HUMGT	P18828 humicola gr
12	184	11.1	1 GUX1_PHACH	P13860 phanerochaete
13	183	11.0	1 GUX1_PENCA	Q06886 neurospora
14	181	10.9	1 GEX1_NEUCR	P38676 neurospora
15	172	10.4	1 GEX1_AGABI	Q00623 agaricus bi
16	159	9.6	1 GUN4_TRIRE	Q14405 trichoderma
17	158.5	9.6	1 GUXC_FUSOX	P46238 fusarium ox
18	156	9.4	1 GUX2_AGABI	Q92400 agaricus bi
19	149.5	9.0	1 GUX1_ASAPC	Q05984 aspergillus
20	143.5	8.6	1 GUN3_HUMIN	Q12624 humicola in
21	137	8.3	1 GUX3_AGABI	P49075 agaricus bi
22	127	7.8	1 GUN2_TRIRE	P07982 trichoderma
23	127	7.7	1 GUNF_FUSOX	P46239 fusarium ox
24	127	7.7	1 GUNB_FUSOX	P46236 fusarium ox
25	127	7.7	1 GUN4_DROME	Q9v6b3 drosophila
26	125.5	7.6	1 LORI_HUMAN	P21490 homo sapien
27	124	7.5	1 GUX2_TRIRE	P07987 trichoderma
28	122	7.3	1 KRUC_SHEEP	P26312 ovis aries
29	120.5	7.3	1 34KD_MYCTU	P71566 mycobacteri
30	120	7.2	1 ZAN_MOUSE	Q08799 mus musculu
31	118.5	7.1	1 PSBP_PORPU	P50272 porphyria pu
32	118.5	7.1	1 VG_DROME	Q26366 drosophila
33	118.5	7.1	1 TROP_HUMAN	Q12816 homo sapien

34	113.5	6.8	2090	1 N214_HUMAN	P35658 homo sapien
35	110.5	6.7	675	1 YWV2_CAEEL	P34504 caenorhabdi
36	109.5	6.6	518	1 FUS_MOUSE	P56959 mus musculu
37	109	6.6	660	1 YHL1_BBY	P03181 epstein-bar
38	107	6.4	967	1 ATSI_RAT	Q9wqg1 rattus norv
39	107	6.4	968	1 ATSI_MOUSE	P97857 mus musculu
40	106.5	6.4	2704	1 G168_PAPPR	P17053 parametium
41	105.5	6.4	452	1 GUX1_CRYPA	Q00548 cryphonectr
42	105	6.3	481	1 LORI_MOUSE	P18165 mus musculu
43	103	6.2	394	1 THID_CLAFS	Q9u14 claviceps f
44	103	6.2	491	1 YK98_MYCTU	Q10707 mycobacteri
45	103	6.2	537	1 SP70_DICDI	P15269 dictyostell

ALIGNMENTS

RESULT 1	GUNK_FUSOX	STANDARD;	PRT;	376 AA.
ID	GUNK_FUSOX			
AC	P45699;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Putative endoglucanase type K precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).			
OS	Fusarium oxysporum.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OC	Hypocreomycetidae; Hypocreales; mitosporic Hypocreales; Fusarium.			
OX	NCBI_Taxid=5507;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=95047531; PubMed=7959045;			
RA	Shepard P.O., Grant F.J., Oort P.T., Sprecher C.A., Foster D.C.,			
RA	Hagen F.S., Upshall A., McKnight G.L., O'Hara P.D.;			
RT	"The use of conserved cellulase family-specific sequences to clone cellulase homologue cDNAs from Fusarium oxysporum.";			
RT	Gene 150:163-167 (1994).			
RL				
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.			
CC	-1- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.			
CC	-1- SIMILARITY: Belongs to cellulase family K (family 45 of glycosyl hydrolases).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).			
CC	EMBL; I29381; AAA65589.1; -.			
DR	HSSP; P43316; ZENG.			
DR	InterPro; IPR009009; Barwin like.			
DR	InterPro; IPR000254; CBD fungal.			
DR	InterPro; IPR000334; Glyco_hydro_45.			
DR	Pfam; PF00734; CBM_1; 1.			
DR	Pfam; PF02015; Glyco_hydro_45; 1.			
DR	SMART; SM00236; FCBD; 1.			
DR	PROSITE; PS00562; CBD_FUNGAL; 1.			
DR	PROSITE; PS01140; GLYCOSYL_HYDROL_P45; 1.			
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.			
FT	SIGNAL	1	18	POTENTIAL.
FT	CHAIN	19	376	PUTATIVE ENDOGLUCANASE TYPE K.
FT	DOMAIN	19	308	CATALYTIC.
FT	DOMAIN	309	338	LINKER.
FT	DOMAIN	339	376	CELLULOSE-BINDING.
FT	ACT SITE	29	29	NUCLEOPHILE (BY SIMILARITY).
FT	ACT SITE	140	140	PROTON DONOR (BY SIMILARITY).
FT	ACT SITE	140	140	PROTON DONOR (BY SIMILARITY).
SO	SEQUENCE	376 AA;	39235 MW;	P43045P62B9F882 CXC64;
Query Match	56.8%;	Score 942.5;	DB 1;	Length 376;

Best Local Similarity 49.5%; Pred. No. 1e-61;
Matches 183; Conservative 36; Mismatches 72; Indels 79; Gaps 6;

QY 9 TTTAAALPLVWASAGSGGSTRYWDCCPKSCAMPKAAVSQPYVACDANFQRLSDFNVOG 68
DB 5 TLLALAGPLAVSAGSGHSTRYWDCCPKSCAGSAAVVAALTCDDNDPISNTNAAV 64
QY 69 GC-NGGSAVSCADQTPANVNDINAFPAATSLAGSSSSMCCACALPFTSGPAGKTV 127
DB 65 GCGGGSAVACTNTPMAVNDLAVGFPAATISGSBASCACALPFTSGPAGKTV 124
QY 128 VQSTGCDLGSNDQDILAMPQGVGI-FNGCSSQFG-GLPAPQVGISSRDQSDFPAPK 186
DB 125 VQSTGCDLGSNDQDILAMPQGVGI-FDQCTSEFGKALGAGAGGIGSSRSCDYPPLAK 184
QY 187 PGQWRPFQWQADNPFTTQVQVCEAFIVASGCKNDSSFPVF----- 232
DB 185 DQGWRFQWQADNPFTTQVQVCEAFIVASGCKNDSSFPVF----- 232
QY 233 -----TPPS-----GNGCTGTPTSPAGSGQTSB 257
DB 245 AKKTTAAAAQPKTKQSAVWQKSTPAAQEPKPADKPFQTDKPAKTPAKTPAQ 304
QY 258 GGGSGCTSOQ-----MAQCGIGF--SGCTTCVSGTTCQ 289
DB 305 PVKPKTKTKVAGTKTRSGCPANMTATAKASVPAVYQCGSKAPVGNALACATGSKCV 364
QY 290 KLANDYISQCL 299
DB 365 KQNEYYSQCV 374

RESULT 2
GUNS_HOMIN
ID GUNS_HOMIN STANDARD; PRT; 213 AA.
AC P43316;
DT 01-NOV-1995 (Ref. 32, Created)
DT 01-NOV-1995 (Ref. 32, Last sequence update)
DT 10-OCT-2003 (Ref. 42, Last annotation update)
DE Endoglucanase V (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DE (Cellulase V) (EG V).
OS Humicola insolens.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
CX NCBI_TaxID=34413;
RN [1]
RP SEQUENCE FROM N.A.
RA Rasmussen G., Mikkelson J.-M., Schuelein M., Packar S.A., Hagen F.,
RA Hjort C.M., Hastrup S.;
RT "A cellulase preparation comprising an endoglucanase enzyme";
RL Patent number WO9117243, 14-NOV-1991.
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RA MEDLINE=93390621; PubMed=837830;
RA Davies G.C., Dodson G.G., Hubbard R.E., Tolley S.P., Dauter Z.,
RA Wilson K.S., Hjort C., Mikkelson J.M., Rasmussen G., Schuelein M.;
RT "Structure and function of endoglucanase V";
RL Nature 365:362-364 (1993).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RA MEDLINE=96101453; PubMed=8519779;
RA Davies G.T., Dodson G.G., Moore M.H., Tolley S.P., Dauter Z.,
RA Wilson K.S., Rasmussen G., Schuelein M.;
RT "Structure determination and refinement of the Humicola insolens
RT endoglucanase V at 1.5-A resolution";
RL Acta Crystallogr. D 52:7-17 (1996).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose, lichenin and cereal beta-D-glucans.

CC -1- SIMILARITY: Belongs to cellulase family K (family 45 of glycosyl
hydrolases).
DB PDB; 2ENG; 08-DEC-96.
DB PDB; 3ENG; 16-JUN-97.
DB PDB; 4ENG; 16-JUN-97.
DB PDB; 1HD5; 03-JUL-01.
DR InterPro: IPR009009; Barvin like.
DR InterPro: IPR00334; Glyco hydro 45.
DR Pfam: PF02015; Glyco hydro 45; 1.
DR PROSITE: PS01140; GLYCOSTYL_HYDROL_F45; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; 3D-structure.
FT ACT SITE 10 10
FT ACT SITE 121 121
FT STRAND 2 8
FT STRAND 12 12
FT HELIX 15 17
FT TURN 19 20
FT STRAND 24 24
FT STRAND 31 31
FT TURN 33 34
FT STRAND 37 37
FT TURN 41 42
FT STRAND 45 45
FT STRAND 46 47
FT TURN 49 50
FT STRAND 53 54
FT TURN 57 58
FT STRAND 62 65
FT TURN 66 67
FT STRAND 68 76
FT TURN 78 79
FT HELIX 82 85
FT TURN 86 87
FT STRAND 89 94
FT HELIX 97 99
FT TURN 100 101
FT STRAND 103 110
FT STRAND 119 123
FT TURN 125 126
FT TURN 130 131
FT HELIX 135 139
FT STRAND 145 145
FT TURN 146 148
FT STRAND 149 149
FT HELIX 153 158
FT HELIX 161 163
FT HELIX 164 171
FT TURN 172 173
FT TURN 175 176
FT STRAND 181 187
FT HELIX 191 197
FT STRAND 201 201
FT TURN 202 203
FT HELIX 204 204
SQ SEQUENCE 213 AA; 22864 MW; 24334301BA3BC804 CRC64;

Query Match 55.5%; Score 921.5; DB 1; Length 213;
Best Local Similarity 74.4%; Pred. No. 1.9e-60;
Matches 157; Conservative 27; Mismatches 26; Indels 1; Gaps 1;

QY 26 GQSTRYWDCCPKSCAMPKAAVSQPYVACDANFQRLSDFNVOGCGN-EGSNVSCADQTPW 84
DB 3 GSTRYWDCCPKSCAMPKAAVSQPYVACDANFQRLSDFNVOGCGN-EGSNVSCADQTPW 62
QY 85 AVNDINAFPAATSLAGSSSSMCCACALPFTSGPAGKTVVQSTGCDLGSNDQD 144
DB 63 AVNDIDALGFPAATSLAGSSSSMCCACALPFTSGPAGKTVVQSTGCDLGSNDQD 122
QY 145 AMPGQGVGI-FNGCSSQFG-GLPAPQVGISSRDQSDFPAPKPGQWRFQWQADNPFT 204
DB 123 NITGGGVGI-FDQCTSEFGKALGAGAGGIGSSRSCDYPPLAKPGQWRFQWQADNPFT 182
QY 205 TFOQVQCEAFIVASGCKNDSSFPVFPTTP 235

FT CHAIN 27 393 ENDOLUCINASE 1.
 FT ACT SITE 34 354 NUCLEOPHILE (BY SIMILARITY).
 FT ACT SITE 152 152 PROTON DONOR (BY SIMILARITY).
 FT DOMAIN 270 385 ALA/GLY/SER-RICH.
 FT CARBOHYD 343 343 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 393 AA; 39594 MW; 65C753C610C6AD3 CRC64;
 Query Match 27.8%; Score 462; DB 1; Length 393;
 Best Local Similarity 30.9%; Pred. No. 9,3e-27;
 Matches 115; Conservative 41; Mismatches 100; Indels 116; Gaps 12;
 Db 11 LAALPLVSAASSSGOSTYMDCKSCAPGKAASQPVYACANQRL---SDPNV 66
 12 LSLSLVHLGDSVRGATRYMDCCASISWEGAPYAVDNCADVTLLDSKDPG 71
 QY QSGNGSAYSCADQTPW--AVNDINAYGPAISIAAGSSSWCNCALFTTSGPYAK 124
 72 QSGNGGNGKFKSCMGPFDEPTPLARFGA--FTTGQSDPDGACVYAF--BHAQK 128
 Db 125 TM----VVQSTSTGDLSSNOFDIAMPGGVGF--NGCSSQFG--GLPGAQYGISR 175
 129 AMKENTLIFQVTVNGGVJGVSQNFQIPGGGIGAFPGYGCFAQWYENASIMGDQYGVKSA 188
 QY 176 DQDSFPAPLXKSGQWRF--WFMADNPTF--TQOYQCAHIVARSGCKNDSSPVF 232
 189 TEGSKLPPLQSGCKMRSEW--GDNPLKSGPRVYCKSLIDRSQCKNDNTISPY 245
 QY 223 T----- 233
 Db 246 SKGVDSANTNAPAYKDRGVCLAGGKSKSAGVYDGSADAGADASGAGNAASQ 305
 QY 224 -----PSSGN-----GCTPTPTSPAGSGTSSG 260
 Db 306 QPEYQCPGSGNDQSGSNQATTAAGSSGSDSGSTANGSGSGAPSDSAAVAPSG 365
 QY 261 SGCTSGKMAQCG 272
 Db 366 SNPQAAGGQGG 377
 RESULT 5
 PABE PENFN STANDARD; PRT; 353 AA.
 AC Q9HEI8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Feruloyl esterase B precursor (BC 3.1.1.73) (ferulic acid esterase B)
 DE (PABE) (Citramoyl esterase).
 GN PABE.
 OS Penicillium funiculosum.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
 OX NCBI_TaxID=28572;
 RN [1] SEQUENCE FROM N.A., SEQUENCE OF 19-68; 90-97; 163-167; 212-241 AND
 RP 253-274, FUNCTION, SUBCELLULAR LOCATION, AND INDUCTION.
 RC STRAIN=IM1 134756;
 RX MEDLINE=20534747; PubMed=11082184;
 RA Kroon P.A., Williamson G., Fish N.M., Archer D.B., Belshaw N.J.;
 RT "A modular esterase from penicillium funiculosum which releases
 RT ferulic acid from plant cell walls and binds crystalline cellulose
 RT contains a carbohydrate binding module.";
 RL Eur. J. Biochem. 267:6740-6752(2000).
 CC -!- FUNCTION: involved in degradation of plant cell walls.
 CC Hydrolyzes the feruloyl-arabinose ester bond in arabinoxylans, and
 CC the feruloyl-galactose and feruloyl-arabinose ester bonds in
 CC pectin. Binds strongly to cellulose.
 CC -!- CATALYTIC ACTIVITY: Feruloyl-polysaccharide + H(2)O = ferulate +
 CC polysaccharide.
 CC -!- ENZYME REGULATION: Inhibited by the specific serine esterase
 CC inhibitor ABBF.
 CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- INDUCTION: Repressed by glucose, probably via the carbon
 CC catabolite repressor protein Crea.
 CC -!- PTM: Glycosylated (Probable).
 CC -!- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AJ291436; CAC1414.1; -.
 DR HSBP; P00725; ZCBH.
 DR GO; GO:0005576; C:extracellular; IDA.
 DR GO; GO:0030248; F:cellulose binding; IDA.
 DR GO; GO:0030600; F:feruloyl esterase activity; IDA.
 DR GO; GO:0016988; P:cell wall catabolism; IDA.
 DR GO; GO:0045490; P:pectin catabolism; IDA.
 DR GO; GO:0045493; P:xylan catabolism; IDA.
 DR InterPro; IPR000254; CBD_fungal.
 DR InterPro; IPR00379; Ser_estrs.
 DR Pfam; PF00734; CBM_1; 1.
 DR Prodom; Pnod1821; CBD_fungal; 1.
 DR SMART; SM00236; fCBD; 1.
 DR PROSITE; PS00562; CBD_FUNGAL; 1.
 DR HydroLase; Serine esterase; Xylan degradation; Glycoprotein; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 353 FERULOYL ESTERASE B.
 FT DOMAIN 19 291 CATALYTIC (PROBABLE).
 FT DOMAIN 292 318 PRO/SER/THR-RICH (LINKER).
 FT DOMAIN 315 353 CELLULOSE-BINDING (PROBABLE).
 FT ACT SITE 136 136 CHARGE RELAY SYSTEM.
 FT CARBOHYD 179 179 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 246 246 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 353 AA; 37333 MW; 4AD08FEDDA9B2D9 CRC64;
 Query Match 14.9%; Score 247; DB 1; Length 353;
 Best Local Similarity 27.5%; Pred. No. 3,6e-11;
 Matches 106; Conservative 28; Mismatches 116; Indels 136; Gaps 21;
 QY 6 VRTTAAALPLVSAASSSGOSTRYWDCCKSCAPGKAASQPVYACANQRL---VYACDANFOR 60
 Db 12 LPLVTLASLITQVNNNGDNPSLQMT-----IYENKLA-SKRALIYVAMHC----- 57
 QY 61 LSDFNWGGNGGSA-----YSCADQ-----TPWAVND-NL--AYGPAISIAAG 102
 Db 58 -----GSAIEYGMVDHSPADQYGLILIPATRDVNCEDAYSSASLTHNGG 106
 QY 103 SES-----SWCCACIALFTTSGPVAKTMVVQSTSTGDLSSNOFDIAMPGGVGFN 155
 Db 107 SDSLIVMYWVYIISTYG-----ADSSKVTMTSSSGALMTN-----VLAGAVDYV- 153
 QY 156 GCSGFGSLPGA-QYGG-----ISSRQXCSFPAPLXKSCW-----RF 193
 Db 154 AAGSFSFSPAPLXKSGADPIMSNQTCGQIQ-HTCQMAAYVHNGYPTGYERL 212
 QY 194 D-WFQADNP-----TFPQVQCAHIVARSGCKR--NDSSEF 230
 Db 213 QMHEGTAADVNTSYADLGEISQWTTIMGLSFTNQNTPL-----SGYTRVYGDGSKFQ 267
 QY 231 VETPSSG-----NGTGTPTSTARSQGTSSGSGTSGQYMAQCG 273
 Db 268 AYSAGVHVPVDPVSWVLMFRTGISTTTTPTTPTTSTPSSTGGCTAAHWAQCG 327
 QY 274 IGFSGCTCVSGTTQCKLNDVYSQCL 299
 Db 328 IGYSCCTCASPYTCQKANDIYSQCL 353
 RESULT 6
 GNS_TRIPE

ID GUN5 TRIE STANDARD; PRT; 242 AA.
AC P43317;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DE (Cellulase V) (EG V).
GN EG-5.
OS Trichoderma reesei (Hypocrea jecorina).
OC Zygomycota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN 11;
RP SEQUENCE FROM N.A.
RC STRAIN=QM9414 / Rut C-30;
RA MEDLINE=95073308; PubMed=7984103;
RA Saloheimo A., Henttinen B., Hoffman A.-M., Telemann O., Penttilae M.,
RT "A novel, small endoglucanase gene, eg1s, from Trichoderma reesei
RT isolated by expression in yeast."
RL Mol. Microbiol. 13:219-228(1994).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -1- SIMILARITY: Belongs to cellulase family K (family 45 of glycosyl
CC hydrolases).
CC -1- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z33381; CAAB3846.1; -
DR PIR; S60143; S60143.
DR HSSP; P00725; 2CBH.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR003112; Expan_endogl.
DR InterPro; IPR00334; Glyco_hydro_45.
DR Pfam; P500734; CBM_1; 1.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR PROSITE; PS00842; EXPANSIN_BG45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 17
FT CHAIN 18 242
FT ACT_SITE 27 27 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 134 134 PROTON DONOR (BY SIMILARITY).
FT DOMAIN 18 182 CATALYTIC.
FT DOMAIN 183 205 PRO/SER-RICH (LINKER).
FT DOMAIN 206 242 CELLULOSE-BINDING (BY SIMILARITY).
FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 213 230 BY SIMILARITY.
FT DISULFID 224 240 BY SIMILARITY.
SQ SEQUENCE 242 AA; 24411 MW; CC033FC51326CT1D CRC64;
Query Match 13.3%; Score 221; DB 1; Length 242;
Best Local Similarity 24.7%; Pred. 1.9e-09;
Matches 78; Conservative 39; Mismatches 101; Indels 98; Gaps 13;

DB 86 APCSSCOTGGAGGSIITMTNLCPNNGAQCPIV--GGTN-----QY 127
QY 171 GISSRDQCSFPAPLPGCCQRRD-WFQN--ADNPFTPGQVCPAEIVANSQ---CKR 223
DB 128 -----GSYHEFDIMAGNEIRFGDVMVWDFEPLACFGQAASDMGTCLCVG 170
QY 224 NDDSPFVFPSPGNGGCTGTPSTIAPSGGQTSPCGGSGCTSGXMMQCGG1PFGSCCTCV 283
DB 171 QQETD-P--TPVLDGMDTSTPSSPPATSSSPGSGGQGT--LYGQCGAGMTPEPTCQ 225
QY 284 SGTCCCKLANDYSSQCL 299
DB 226 APGTCKYQNMYSQCL 241
RESULT 7
ID GUN1 TRIE STANDARD; PRT; 513 AA.
AC P00725;
DT 21-UTR-1986 (Rel. 01, Created)
DT 21-UTR-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellulohydrolase I)
DE (CBHI) (1,4-beta-cellulohydrolase).
GN CBHI.
OS Trichoderma reesei (Hypocrea jecorina), and
OS Trichoderma koningi.
OC Zygomycota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453; 55202;
RN 11;
RP SEQUENCE FROM N.A.
RC SPECIES=T.reesei; STRAIN=L27;
RA Shoemaker S., Schweickart V., Ladner M., Gelfand D., Kwok S.,
RT Myambo K., Imis M.;
RT "Molecular cloning of exo-cellulohydrolase I derived from Trichoderma
RT reesei strain L27."
RL Biotechnology 1:691-696(1983).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=T.koningii; STRAIN=G-39;
RX MEDLINE=94100788; PubMed=7764306;
RA Way T.T., Hsueh T.H., Huang L.;
RT "Molecular cloning and sequence analysis of the cellulohydrolase I
RT gene from Trichoderma koningii G-39."
RL Curr. Microbiol. 28:31-39(1994).
RN [3]
RP ACTIVE SITE.
RC SPECIES=T.reesei;
RA Tome P., Claysens M.;
RT "Identification of a functionally important carboxyl group in
RT cellobiohydrolase I from Trichoderma reesei."
RL FEMS Lett. 243:239-243(1989).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.81 ANGSTROMS) OF 18-452.
RX MEDLINE=94310436; PubMed=8036495;
RA Dyrne C., Staehberg U., Reinikainen T., Ruohonen L., Petersson G.,
RT Knowles J.K.C., Teeri T.T., Jones T.A.;
RT "The three-dimensional crystal structure of the catalytic core of
RT cellobiohydrolase I from Trichoderma reesei."
RL Science 265:524-528(1994).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 18-452.
RC SPECIES=T.reesei; STRAIN=QM9414 / Rut C-30;
RX MEDLINE=98128795; PubMed=9466911;
RA Dyrne C., Staehberg U., Teeri T.T., Jones T.A.;
RT "High-resolution crystal structures reveal how a cellulose chain is
RT bound in the 50 A long tunnel of cellobiohydrolase I from Trichoderma
RT reesei."
RL J. Mol. Biol. 275:309-325(1998).
RN [6]
RP STRUCTURE BY NMR OF 478-513.

RC SPECIES=I.reesei;
RX MEDLINE=90057416; PubMed=2554967;
RA Krawiec P.J., Clore G.M., Nilges M., Jones T.A., Petersson G.,
RA Krawiec J., Gronenborn A.M.;
RT "Determination of the three-dimensional solution structure of the C-
terminal domain of cellobiohydrolase I from Trichoderma reesei. A
RT study using nuclear magnetic resonance and hybrid distance
RT geometry-dynamical simulated annealing.";
RL Biochemistry 28:7241-7257(1989).
RN [7].
RP STRUCTURE BY NMR OF 478-513.
RC SPECIES=I.reesei;
RX MEDLINE=91194052; PubMed=3041630;
RA Mattinen M.L., Konttinen M., Kerovuo J., Linder M., Annala A.,
RA Lindberg G., Reinikainen T., Drakenberg T.;
RT "Three-dimensional structures of three engineered cellulase-binding
RT domains of cellobiohydrolase I from Trichoderma reesei.";
RL Protein Sci. 6:294-303(1997).
CC -I- FUNCTION: The biological conversion of cellulose to glucose
CC generally requires three types of hydrolytic enzymes: (1)
CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)
CC Exocellobiohydrolases that cut the disaccharide cellobiose from
CC the nonreducing end of the cellulose polymer chain; (3) Beta-1,4-
CC glucosidases which hydrolyze the cellobiose and other short cello-
CC oligosaccharides to glucose.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- MISCELLANEOUS: T.reesei produces two different
CC exocellobiohydrolases. They are unique in that they can hydrolyze
CC crystalline cellulose in the absence of endoglucanases.
CC -I- SIMILARITY: Belongs to cellulase family C (family 7 of glycosyl
CC hydrolases).
CC -I- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X69976; CAA49596.1; -.
DR PIR: A03302; EUTCI-
DR PDB: 1CEU; 01-NOV-94.
DR PDB: 2CEU; 12-MAR-97.
DR PDB: 3CEU; 12-MAR-97.
DR PDB: 4CEU; 12-MAR-97.
DR PDB: 5CEU; 24-DEC-97.
DR PDB: 6CEU; 24-DEC-97.
DR PDB: 7CEU; 24-DEC-97.
DR PDB: 1CBH; 15-JAN-90.
DR PDB: 2CBH; 15-JAN-90.
DR PDB: 1AZ6; 08-APR-98.
DR PDB: 1AZH; 08-APR-98.
DR PDB: 1AZJ; 29-APR-98.
DR PDB: 1AZK; 29-APR-98.
DR PDB: 1DY4; 18-DEC-00.
DR PDB: 1BGN; 16-MAY-01.
DR PDB: 8CEU; 25-FEB-98.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR008985; ConA_like_1ec.g1.
DR InterPro: IPR001722; Glyco_hydro_7.
DR Pfam: PF00734; CBM_1.1.
DR Pfam: PF00840; Glyco_hydro_7; 1.
DR ProDom: PD001821; CBD_fungal; 1.
DR ProDom: PD186135; Glyco_hydro_7; 1.
DR SMART: SMO0236; FCBM; 1.
DR PROSITE: PS00562; CBD_FUNGAL; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal;
KW 3D-structure; Pyrrolidone carboxylic acid.

FT SIGNAL 1 17
FT CHAIN 18 513
FT DOMAIN 18 453
FT DOMAIN 454 477
FT DOMAIN 478 513
FT ACT_SITE 143 143
FT ACT_SITE 229 229
FT ACT_SITE 234 234
FT MOD_RES 18 18
FT CARBOHYD 62 62
FT CARBOHYD 81 81
FT CARBOHYD 287 287
FT CARBOHYD 401 401
FT DISULFID 21 89
FT DISULFID 36 42
FT DISULFID 67 88
FT DISULFID 78 84
FT DISULFID 155 414
FT DISULFID 189 227
FT DISULFID 193 226
FT DISULFID 247 273
FT DISULFID 255 260
FT DISULFID 278 348
FT DISULFID 485 502
FT DISULFID 496 512
FT STRAND 19 20
FT STRAND 30 36
FT TURN 38 39
FT STRAND 42 51
FT HELIX 53 55
FT STRAND 58 60
FT TURN 61 62
FT STRAND 66 66
FT STRAND 68 69
FT TURN 70 71
FT STRAND 72 73
FT TURN 75 77
FT HELIX 81 87
FT STRAND 88 90
FT TURN 95 99
FT STRAND 101 104
FT TURN 105 106
FT STRAND 107 111
FT STRAND 113 115
FT STRAND 119 121
FT STRAND 123 130
FT TURN 131 132
FT STRAND 133 134
FT STRAND 136 137
FT TURN 140 141
FT STRAND 142 149
FT TURN 151 152
FT TURN 155 156
FT STRAND 157 164
FT TURN 168 174
FT TURN 176 177
FT HELIX 182 184
FT TURN 185 185
FT STRAND 190 190
FT TURN 191 192
FT STRAND 199 200
FT TURN 201 202
FT STRAND 203 204
FT TURN 207 208
FT STRAND 210 211
FT TURN 215 217
FT STRAND 219 220
FT STRAND 223 226
FT STRAND 229 235
FT STRAND 240 245
FT STRAND 247 247
FT STRAND 253 256

EXOGUCANASE I.
CATALYTIC.
LINER
CELLULOSE-BINDING.
PROBABLE.
NUCLEOPHILE.
PROTON DONOR.
PYRROLIDONE CARBOXYLIC ACID.
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).

FT HELIX 367 373
 FT TURN 375 376
 FT STRAND 378 387
 FT TURN 389 392
 SQ SEQUENCE 459 AA; 48208 MW; D235A256P8C8CB9 CRCE4;

Query Match 12.9%; Score 214; DB 1; Length 459;
 Best Local Similarity 27.0%; Pred. No. 1.2e-08;
 Matches 108; Conservative 31; Mismatches 101; Indels 160; Gaps 27;

QY 19 ASAGSGGQS---TRYWDCCKPSGAPGKAASQPYVACDA-----NFORLSDENV 66
 DB 101 ASGVTSSSSITMXYQ---MPSSS-GGYSVSPRLYLDSGKIVMLKNGEIS-PDV 154
 QY 67 Q-----SCGSGSASVSCADQTP---MAVNDNLAYGFA 95
 DB 155 DLALPCGENSLYLQOMENOGANQVITAGAVNGGY-CLAQGVQVQW-NQTL----- 207
 QY 96 ATSIAGSESSKCC-----AC-----YALITFS-GPV 121
 DB 208 -----NTSHGGCCNEMDILBGNSPANLTPHSCATACDSAGCGFNPYSGYKSYGP- 262
 QY 122 AGKMWVQST-----STGGDLS-----NOFDI-AMPGGVSIFENGCSQ 160
 DB 262 -GVVDSKRTFTITQENTDNGSPGNVSVITKYGQNGVDLPASQSGDT--ISSCF8- 317
 QY 161 FGLGAGQYGGISRDQCSFPAPLKPCCQWRD-WFONAD-NPTTFQOVQC----- 211
 DB 318 -----ASAYGGLATMGKA-----LSSGMVIVFSIWNDSQYMWLDSGNAGCSSTBGN 366
 QY 212 PAIYARSGCKNDSSFPVTPPGGNGGNGTPTTA-----PSGGT-----SSGG 259
 DB 367 PSNIIA-----NNPNTHVFSNIRMGDIGS-TTNSIAPPPPASSTFTSTRSSSTYS 419
 QY 260 GSAGCTSGKMAQCGIGFSGCTTCGTLNDYSSQCL 299
 DB 420 SPSCQTHMGQCGSIGSGCKTCTGTTTCQYNDYSSQCL 459

RESULT 10
 GUN1_TRILIO STANDARD; PRT; 463 AA.
 ID GUN1_TRILIO
 AC 012714;
 DT 01-NOV-1997 (Rel. 35, Created);
 DT 01-NOV-1997 (Rel. 35, Last sequence update);
 DT 10-OCT-2003 (Rel. 42, Last annotation update);
 DB Endoglucanase EG-1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase).
 GN EGI1.
 OS Trichoderma longibrachiatum.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreomycetidae; Hypocreales; mitosporic Hypocreales; Trichoderma.
 OX NCBI_TaxID=5548;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=CECT 2606;
 RA Perez-Gonzalez J. A.;
 RL Submitted (NOV-1992) to the EMBL/GenBank/DDBJ databases.
 CC -1- FUNCTION: The biological conversion of cellulose to glucose
 CC generally requires three types of hydrolytic enzymes: (1)
 CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)
 CC Exocellulohydrolases that cut the disaccharide celllobiose
 CC from the nonreducing end of the cellulose polymer chain; (3)
 CC Beta-1,4-glucosidases which hydrolyze the celllobiose and other
 CC short cello-oligosaccharides to glucose.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.
 CC -1- STRUCTURAL LOCATION: Secreted.
 CC -1- SIMILARITY: Belongs to cellulase family C (family 7 of glycoyl
 CC Hydrolases).
 CC -1- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL; X60652; CAA43059.1; -;
 DR HSRP; A48375; A48375.
 DR HSRP; P07981; IEG1.
 DR InterPro; IPR000254; CBD_fungal.
 DR InterPro; IPR008985; ConA_like_1ec_g1.
 DR InterPro; IPR001722; Glyco_hydro_7.
 DR Pfam; PF00734; CBM_1; 1.
 DR Pfam; PF00840; Glyco_hydro_7; 1.
 DR ProDom; PD186135; Glyco_hydro_7; 1.
 DR SMART; SM00236; fCBD; 1.
 DR PROSITE; PS00562; CBD_FUNGAL; 1.
 KM Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 463
 FT DOMAIN 23 397
 FT DOMAIN 402 427
 FT DOMAIN 428 463
 FT ACT_SITE 218 218
 FT ACT_SITE 223 223
 FT DISULFID 435 452
 FT DISULFID 446 462
 FT CARBOHYD 78 78
 FT CARBOHYD 164 164
 FT CARBOHYD 204 204
 FT CARBOHYD 208 208
 FT CARBOHYD 394 394
 SQ SEQUENCE 463 AA; 48337 MW; B3AC3DPD3ADZELC CRCE4;

Query Match 12.2%; Score 203; DB 1; Length 463;
 Best Local Similarity 28.4%; Pred. No. 7.5e-06;
 Matches 86; Conservative 22; Mismatches 101; Indels 94; Gaps 18;

QY 68 SGCGSGSASVSCADQTP---MAVNDNL---AYGFAAT-SIAGSESSKCCACVATPTSG 119
 DB 184 AGAYNSGCV-COAQCPVQVQW-NQTLNTSGGFCNEMDILBGNSPANLTPHSCATAC 241
 QY 120 PVAG-----KTMVY-----QSTSTGDLGS-----NOFD 143
 DB 242 DSACGCFNPYSGGYNYPFGDVTWTSKFTITITQFNTDNGSPGNVSVITKYGQNGVD 301
 QY 144 I--AMPQGVGIFNCGSSQFGLGAGQYGGISRDQCSFPAPLKPCCQWRD-WFONAD 200
 DB 302 IPSAKPGSDT--ISSCF8-----ASAYGGLATMGKA-----LSSGMVIVFSIWNDSQ 347
 QY 201 -----NPTTFQOVQCPAIVARSCKRNDSSFPVTPPGSN--- 239
 DB 348 YMWMLDSGRAGPCSSTEGNPNIL--ANNPQTHVIVYSNIRMGDIGS--TTNSGSGNPP 402
 QY 240 ---GCTETPTSTAPSSGQTPSGGSGCTSGKMAQCGIGFSGCTTCVSGTTCQKNDYSS 296
 DB 403 PPPPASSTFTSTRSSSTTS--SSPSCQTHMGQCGIGYGTCKTCTGTTTCQYNDYSS 460
 QY 297 QCL 299
 DB 461 QCL 463

RESULT 11
 GUN1_HUMGE STANDARD; PRT; 525 AA.
 ID GUN1_HUMGE
 AC P15828;
 DT 01-APR-1990 (Rel. 14, Created);
 DT 01-FEB-1996 (Rel. 33, Last sequence update);
 DT 10-OCT-2003 (Rel. 42, Last annotation update);
 DB Exoglucanase I precursor (EC 3.2.1.91) (Exocellulohydrolase I) (1,4-
 DB beta-cellobiohydrolase) (Beta-glucanase/cellobiohydrolase).

GN Humicola grisea var. thermoides.
 OS Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
 OX NCBI_TaxID=5528;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90175006; PubMed=2308855;
 RA de Oliveira Alzevedo M., Radford A.;
 RT "Sequence of chb-1 gene of Humicola grisea var. thermoides.";
 RL Nucleic Acids Res. 18:668-668(1990).
 CC -1- FUNCTION: The biological conversion of cellulose to glucose
 CC generally requires three types of hydrolytic enzymes: (1)
 CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)
 CC Exocellulohydrolases that cut the disaccharide cellobiose from
 CC the nonreducing end of the cellulose polymer chain; (3) Beta-1,4-
 CC glucosidases which hydrolyze the cellobiose and other short cello-
 CC oligosaccharides to glucose.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellobiose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SIMILARITY: Belongs to cellulase family C (family 7 of glycosyl
 CC hydrolases).
 CC -1- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X17258; CA35159.1; -.
 DR PIR; S38794; S38794.
 DR HSSP; P00725; 8CBD.
 DR InterPro; IPR000254; CBD_fungal.
 DR InterPro; IPR008985; Coma_like_1ec_g1.
 DR InterPro; IPR001722; Glyco_hydro_7.
 DR Pfam; PF00734; Cbm_1; 1.
 DR Pfam; PF00840; Glyco_hydro_7; 1.
 DR ProDom; PD001821; CBD_fungal; 1.
 DR ProDom; PD186135; Glyco_hydro_7; 1.
 DR SMART; SMO0236; fCBD; 1.
 DR PROSITE; PS00562; CBD_FUNGAL; 1.
 KM Cellulose degradation: Hydrolase; Glycosidase; Glycoprotein; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 525
 FT DOMAIN 19 467
 FT DOMAIN 468 489
 FT DOMAIN 490 525
 FT ACT SITE 231 231
 FT ACT SITE 236 236
 FT CARBOHYD 289 289
 FT DISULFID 497 514
 FT DISULFID 508 524
 SQ SEQUENCE 525 AA; 55693 MW; A6684D4C8E81B090 CRC64;
 Query Match 11.2%; Score 185; DB 1; Length 525;
 Best Local Similarity 24.9%; Pred. No. 1.7e-06;
 Matches 97; Conservative 33; Mismatches 136; Indels 124; Gaps 23;
 QY 11 LAALPLVASASGSGSTRVMDCKKSCAMPK-KAASQPVYACDAPFQLSPF----- 64
 DB 159 LAGALVFSVMDADGG--LSRY-----FGKKAAXKGYCYDACCPRDIKFINGRA 206
 QY 65 NVDSGNG-----GGSNIS--CAQOTPMANVMILAYGR--NTSIAGSS--ESSWCAC 111
 DB 207 NIEGWTGSTNDPNNAGRGYCCCESEMDIWEAN--NMAFAFEHPCITIIQSRCBDSGCGT 265
 QY 112 YALFTPSG-----FVAGKTVVQST-----STGGDGS--NGF 142
 DB 266 YSNERYAGVCPDGDGDNRSYRQGNKTFYKGNKIVDTIKITVYQFLDANGDLGEIKKF 325

QY 143 -----DIAMPG--GGVGFNGCSSO---FGGLPG-AOYGGISNRDQCSFPAFL 185
 DB 326 YVQDQKIIPISESTTTEGEGNSITDWCORQKARFDDIDPRKNGKQKAKLAPKML 385
 QY 186 KPGCWRFDFQFONADNPTF---TEQVQCPAEIVARSCKEN---DDSEFPV----- 231
 DB 386 VMS-----WDIASMMMLDSTF-----PYDAKGPAGAEAGCAPTISGVPAVEAEAPN 435
 QY 232 -----FTP-----DSGNGGTG-----TPSTAPSGQTSFGGSGSCTSQMA 263
 DB 436 SNVFSNRIRGPIGSTVAGLPAGAGNGNCGNDPPEPTTTSAPATTTASAPKRWQ 495
 QY 270 QCGGIFGSGCTGVSCTGCKXNDYISQCL 299
 DB 496 QCGGIFGTPQCEBYICTKNDYISQCL 525
 RESULT 12
 ID GIXL PHACH STANDARD; PRT; 516 AA.
 AC P13860;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellulohydrolase I)
 DE (1,4-beta-cellobiohydrolase).
 GN Phanerochaete chrysosporium.
 OS Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Aphyllophorales; Corticiaceae; Phanerochaete.
 OX NCBI_TaxID=5306;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 34541 / ME446;
 RX MEDLINE=89232732; PubMed=3246351;
 RA Sims P.F.G., James C., Broda P.;
 RT "The identification, molecular cloning and characterisation of a gene
 RT from Phanerochaete chrysosporium that shows strong homology to the
 RT exo-cellobiohydrolase I gene from Trichoderma reesei.";
 RL Gene 74:411-422(1988).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 34541 / ME446;
 RX MEDLINE=9435641; PubMed=8057846;
 RA Sims P.F.G., Soares-Felipe M.S., Wang Q., Gent M.E., Tempelaar C.,
 RA Broda P.;
 RT "Differential expression of multiple exo-cellobiohydrolase I-like
 RT genes in the lignin-degrading fungus Phanerochaete chrysosporium.";
 RL Mol. Microbiol. 12:209-216(1994).
 CC -1- FUNCTION: The biological conversion of cellulose to glucose
 CC generally requires three types of hydrolytic enzymes: (1)
 CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)
 CC Exocellulohydrolases that cut the disaccharide cellobiose
 CC from the nonreducing end of the cellulose polymer chain; (3)
 CC Beta-1,4-glucosidases which hydrolyze the cellobiose and other
 CC short cello-oligosaccharides to glucose.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellobiose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Belongs to cellulase family C (family 7 of glycosyl
 CC hydrolases).
 CC -1- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M22220; AAB46373.1; -.

DR EMBL: Z22528; CAA80253.1; -
 DR PIR: J00083; J00083.
 DR HSSP: P00725; 2CBH.
 DR InterPro: IPR000254; CBD_fungal.
 DR InterPro: IPR008985; ConA_like_1ec.gi.
 DR InterPro: IPR001722; Glyco_hydro_7.
 DR Pfam: PF00734; CBM 1; 1.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR ProDom: PD01821; CBD_fungal; 1.
 DR ProDom: PD186135; Glyco_hydro_7; 1.
 DR SMART: SM00236; CBD; 1.
 DR ProSITE: PS00562; CBD_FUNGAL; 1.
 XW Cellulose degradation: Hydrolyase; Glycosidase; Glycoprotein; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 516
 FT DOMAIN ? 449
 FT DOMAIN 450 480
 FT ACT_SITE 481 516
 FT ACT_SITE 225 225
 FT ACT_SITE 230 230
 FT CARBOHYD 208 208
 FT CARBOHYD 326 326
 FT CARBOHYD 442 442
 FT DISULFID 488 505
 FT DISULFID 499 515
 FT CONFLICT 27 28
 FT CONFLICT 30 31
 FT CONFLICT 31 31
 SQ SEQUENCE 516 AA; 54857 MW; 1C7C3D338CE1B72 CRC64;
 Query Match 11.1%; Score 184; DB 1; Length 516;
 Best Local Similarity 24.7%; Pred. No. 2e-06;
 Matches 86; Conservative 31; Mismatches 113; Indels 112; Gaps 20;
 QY 20 SAASGSGSTRYWDCKPCSPAMPK--AAVSQPYACANFQRLSDFNVSQGNCG-- 73
 DB 211 SAAAGT---NYGTCTEMDMYMANNDAAATP-HPCTT-----NAQRCSSGSDCTR 258
 QY 74 -----SAVCAQDTPWAVNDLAAFGPAISISAGSESSWCACALTFPSGPV 121
 DB 259 DTGLCAQDCDFNSFRMGDT-----FLKGLVYDTSKPFVTYQFTT-NDGTS 306
 QY 122 AG-----KTYVVGSTGTGDLGSGNDFIAMFG-AGVGIF--NGCSSQ---EGGIP-GAQY 169
 DB 307 AGTLTEIRLTYVN---GVYONS-SYKIGIDPVMSITDIFGCGQKTAEGDYNFAQH 361
 QY 170 GGISRDQC-----DSFAPLPKPCQMRPEWQ-----NADNPTTFQVQC- 211
 DB 362 GELKYGEALRTGVYALTS-WDZYANML-----WLDQVYPTNKDPSFPGVARGTCA 413
 QY 212 -----PAEIVARSGCKRNDSSFPFTPSGNGCT-----GPTSTAP 250
 DB 414 TTSGVPAQIEAQSS-----FNAYVFSNIRKGDMLTYGTGVSSSSVSSHSSTJSSSH 467
 QY 251 GSGQTSFGGSGCTSCKMAQCGGIGFSGCTTCEGTTCCGLNDYSQC 298
 DB 468 SSSSTFGTPTGVTTPVQMGCGSIGTGSITTCAPFYCHLVNPIYSQC 515
 RESULT 3
 GUXI_PENCA
 ID GUXI_PENCA STANDARD; PRT; 537 AA.
 AC 006886; 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DB Exoglucanase I precursor (EC 3.2.1.91) (Exocellulohydrolase 1)
 DB (1,4-beta-cellulohydrolase).
 GN CBH1.
 OS Penicillium janthinellum (Penicillium vitale).
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 CC Eubacterales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
 CX NCBI_Taxid=5079;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=C41;
 RX MEDLINE=93178976; PubMed=8440481;
 RA Koch A., Weigel C.T.O., Schult G.;
 RT "Cloning, sequencing, and heterologous expression of a cellulase-
 encoding cDNA (cbh1) from Penicillium janthinellum";
 RL Gene 124:57-65(1993).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 in cellulose and cellobiose, releasing cellobiose from the non-
 reducing ends of the chains.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Belongs to cellulase family C (family 7 of glycosyl
 hydrolases).
 CC -1- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: X59054; CAA1780.1; -
 DR PIR: J00150; J00150.
 DR HSSP: P00725; 8CEU.
 DR InterPro: IPR000254; CBD_fungal.
 DR InterPro: IPR008985; ConA_like_1ec.gi.
 DR InterPro: IPR001722; Glyco_hydro_7.
 DR Pfam: PF00734; CBM 1; 1.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR ProDom: PD001821; CBD_fungal; 1.
 DR ProDom: PD186135; Glyco_hydro_7; 1.
 DR SMART: SM00236; CBD; 1.
 DR ProSITE: PS00562; CBD_FUNGAL; 1.
 KV Cellulose degradation: Hydrolyase; Glycosidase; Glycoprotein; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 537
 FT DOMAIN 19 453
 FT DOMAIN 454 477
 FT DOMAIN 478 537
 FT ACT_SITE 235 235
 FT ACT_SITE 240 240
 FT ACT_SITE 136 136
 FT CARBOHYD 414 414
 FT CARBOHYD 456 456
 FT CARBOHYD 509 526
 FT DISULFID 520 536
 FT DISULFID 537 536
 SQ SEQUENCE 537 AA; 56842 MW; A6B9C6EB73F17FE4 CRC64;
 Query Match 11.0%; Score 183; DB 1; Length 537;
 Best Local Similarity 23.4%; Pred. No. 2.5e-06;
 Matches 79; Conservative 30; Mismatches 99; Indels 130; Gaps 14;
 QY 24 GSGQSTRYWDCKPCSPAMPKAAVSQPYACANFQRLSDFNVSQGNCGSAYACADQTP 83
 DB 268 GTYSTDRYGTCDPD-----CDNPFYRGVYNYF-----GPGETIDTFRSP 308
 QY 84 WAV-----NDNLAYFPA-----TSIAGS-SSWCACALTF 116
 DB 309 FTVVQELTMDGTSTLSEIKRFYVQSKVIGNPSTIVGSGNSITDSWCA----- 362
 QY 117 TSGPAGKIMVVGSTGTGDLGSGNDFI--AMPGGVGIFNGC----- 157
 DB 363 -----QKSAFSD--TNEFSKKGNAQKAGLADGVYLVMSLMDHSDMLML 407
 QY 158 -----SFGGLPQAQYGGISRDQCSFAPLPKPCQMRPEWQNDNPTTFQVQCP 212
 DB 408 DSTYVTAISTTPGAK-----KGTCDISRRP-----NTVESTY-----P 441
 QY 213 AEIVARSGCKRNDSSFPFTPSGNGGIGTGTSTAPRSCQ-----TSRGGGS 261
 DB 442 NAYVYFSNIRKGPLNS--TFTGGTSSSITTTTSKSTSTSSSKTTTITVTTTSSSS 499

DR PROSITE; PS00562; CBD_FUNGAL; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 29
FT CHAIN 30 320
FT DOMAIN 30 261
FT DOMAIN 252 285
FT DOMAIN 286 320
FT DISULFID 232 309
FT DISULFID 323 319
FT CARBOHYD 163 163
SQ SEQUENCE 320 AA; 33754 MW; 60E2C8C80895CA2B CRC64;
N-LINKED (GLCNAC...) (POTENTIAL).
BY SIMILARITY.
CELLULOSE-BINDING (POTENTIAL).
CELLULOSE-GROWTH-SPECIFIC PROTEIN.
CATALYTIC (POTENTIAL).
POTENTIAL.
Query Match 10.4%; Score 172; DB 1; Length 320;
Best Local Similarity 24.8%; Pred. No. 9.2e-06;
Matches 85; Conservative 36; Mismatches 132; Indels 90; Gaps 15;
QY 10 TLAALPLVNSAASGGSGOSTRY-WDCC-----KPSGAMPGRKAAVSQPVYACDANFQRLSD 63
DB 15 TFLALGLFAKVVQAHGVIGYSMDGTWYBGWHPYNTPVGCTSIERPW---ATFDPIMD 70
QY 64 FNVQS-GGNGGSAYSQADQTPMAVNDNLAYPAATSIAGSSSSWCCACTALTFTSGPYA 122
DB 71 ATASTVGCNN-----DGNP-GPQULTATVAAGTALTAYNQVWPHRYGPMITTYLGKCP 122
QY 123 GKTWVVGSTST-----GGDLG-----SNQPDIAIPGGVGIFNGCS 158
DB 123 GSSCGVNTNLSKMFIDENGLSTGVKXGWSGKIDQNNMTTIP----- 171
QY 159 SQFGGLPGAQYGGISSRQCDSPFPAFLKPGCQMRFDWFQMANDPFTFPQVQCP--AEIV 216
DB 172 ---STVPSGAYMIFETIALHSLPAQIYPECA-----QLITGGGNRAPTSSEIV 218
QY 217 ARSGGKRD-----DSSFPVTPP---SGANG--TGTPTSTAPSGGQTS 256
DB 219 SPPGYSNSDPGLTVNLVYQEMMDFTYIVGPPLYSGGNGSSPTTPTTPTTPTTSP 278
QY 257 PGSGSGCTSKMAQCGAIGFSGCTCVSGTTCOKLNDYYSCCL 299
DB 279 PTSTPG-TIPYGGCGSIGMTGTGTCVAPYQCKVINDYSGCL 320

Search completed: July 7, 2004, 08:53:23
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CV protein - protein search, using sw model

Run on: July 7, 2004, 08:47:55 ; Search time 45 Seconds

(without alignments)
2096.443 Million cell updates/sec

Title: US-10-007-521-12

Sequence: 1659

1 MRSPTVTRTAAALPLVAS.....TTTCSTGTCCKNDYRSQL 299

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_Archaea:*
2: SP_Bacteria:*
3: SP_Fungi:*
4: SP_Human:*
5: SP_Invertebrate:*
6: SP_Mammal:*
7: SP_Rbc:*
8: SP_Organelle:*
9: SP_Phage:*
10: SP_Plant:*
11: SP_Rodent:*
12: SP_Virus:*
13: SP_Vertebrate:*
14: SP_Unclassified:*
15: SP_Virus:*
16: SP_Bacteriophage:*
17: SP_Archaea:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1272	76.7	293	3	Q87201	Q87201 neurospora
2	1170	70.5	305	3	Q93782	Q93782 humicola gr
3	1039	62.6	235	3	Q800X8	Q93088 meliocarpa
4	757.5	45.7	366	3	Q80110	Q93110 rhizopus or
5	756.5	45.6	338	3	Q80112	Q93112 rhizopus or
6	745	44.9	360	3	Q80111	Q93111 rhizopus or
7	724	43.6	219	14	Q93992	Q93992 unclassified
8	724	43.6	220	14	Q93984	Q93984 unclassified
9	709	42.7	219	14	Q93991	Q93991 unclassified
10	706.5	42.6	220	14	Q93983	Q93983 unclassified
11	705	42.5	219	14	Q93995	Q93995 unclassified
12	703	42.4	219	14	Q93986	Q93986 unclassified
13	700	42.2	219	14	Q93987	Q93987 unclassified
14	698.5	42.1	227	3	Q93783	Q93783 humicola gr
15	698	42.1	219	14	Q93990	Q93990 unclassified
16	697.5	42.0	221	14	Q93994	Q93994 unclassified

17	697	42.0	219	14	Q93988	Q93988 unclassified
18	697	42.0	219	14	Q93989	Q93989 unclassified
19	693	41.8	220	14	Q93996	Q93996 unclassified
20	689.5	41.6	220	14	Q93993	Q93993 unclassified
21	680.5	41.0	219	14	Q93985	Q93985 unclassified
22	632	38.1	219	5	Q7Y278	Q7Y278 mastoetermes
23	631	38.0	410	3	Q9P868	Q9P868 pironces e
24	625	37.7	219	5	Q7Y277	Q7Y277 mastoetermes
25	617.5	37.2	220	5	Q7Y279	Q7Y279 mastoetermes
26	613.5	37.0	220	5	Q7Y280	Q7Y280 mastoetermes
27	585.5	35.3	271	3	Q9UPV3	Q9UPV3 alternaria
28	543.5	32.8	237	5	Q81815	Q81815 alternaria ger
29	543.5	32.8	242	5	Q97401	Q97401 phaeodon goc
30	360.5	21.7	112	14	Q93999	Q93999 unclassified
31	348.5	21.0	112	14	Q939A0	Q939A0 unclassified
32	347.5	20.9	112	14	Q939H8	Q939H8 unclassified
33	338.5	20.4	112	14	Q939A2	Q939A2 unclassified
34	338.5	20.4	112	14	Q939A1	Q939A1 unclassified
35	330	19.9	111	14	Q939H7	Q939H7 unclassified
36	221	13.3	247	3	Q7Z7X0	Q7Z7X0 trichoderma
37	215.5	13.0	459	3	Q7Z7X3	Q7Z7X3 trichoderma
38	215	13.0	514	3	Q93832	Q93832 trichoderma
39	200	12.1	505	3	Q9P8P3	Q9P8P3 trichoderma
40	191	11.5	525	3	Q12621	Q12621 humicola gr
41	188.5	11.4	302	3	Q99034	Q99034 trichoderma
42	188	11.3	293	3	Q871E8	Q871E8 neurospora
43	187.5	11.3	316	3	Q7Z299	Q7Z299 trichoderma
44	183	11.0	523	3	Q9Y723	Q9Y723 iprex lacte
45	181.5	10.9	529	3	Q8WZJ4	Q8WZJ4 penicillium

ALIGNMENTS

RESULT 1	
Q87201	PRELIMINARY; PRT; 293 AA.
AC Q87201	
DT 01-JUN-2003 (TrEMBLrel. 24, Created)	
DT 01-OCT-2003 (TrEMBLrel. 24, Last sequence update)	
DE Hypothetical protein B19A17.010.	
GN B19A17.010.	
OS Neurospora crassa.	
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.	
OX NCBI_TaxID=5141;	
RN [1]	
RP SEQUENCE FROM N.A.	
RA Schulte U., Aign V., Hoheisel U., Brandt P., Fartmann B., Holland R.,	
RA Nyakatura G., Mewes H.W., Mannhaupt G.;	
RU Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.	
RN [2]	
RP SEQUENCE FROM N.A.	
RA German Neurospora genome project;	
RU Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.	
DR EMBL; BX284763; CAD70529.1;	
DR GO; GO:0005576; C:extracellular; IEA.	
DR GO; GO:0008810; F:cellulase activity; IEA.	
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.	
DR InterPro; IPR009009; Barwin like.	
DR InterPro; IPR000254; CBD_fungal.	
DR InterPro; IPR000354; Glyco_hydro_45.	
DR Pfam; PF00734; CBM_1; 1.	
DR Pfam; PF02015; Glyco_hydro_45; 1.	
DR ProDom; PD001821; CBD_fungal; 1.	
DR SMART; SMO0236; fCBD; 1.	
DR PROSITE; PS00552; CBD_FUNGAL; 1.	
DR PROSITE; PS01440; GLYCOSYL_HYDROL_F45; 1.	
KW Hypothetical protein.	
SQ SEQUENCE 293 AA; 30261 MW; 27096407F349877A CRC64;	
Query Match 76.7%; Score 1272; DB 3; Length 293;	

Best Local Similarity 74.8%; Pred. No. 1,5e-93;
Matches 223; Conservative 30; Mismatches 39; Indels 6; Gaps 2;

QY 1 MRSTPVLRITLAAALPLVSAASGSGSTRYWDCKRSCAMPKAAVSQPYVACDANFOR 60
DB 1 MRSTPVLRITLAAALPLVSAASGSGSTRYWDCKRSCAMPKAAVSQPYVACDANFOR 60
QY 61 LSDENVVSGGNGSAYSCADOTPMVANDNLAYGPAATSTAGSESSWCACATLTFTSG 120
DB 61 LSDENVVSGGNGSAYSCADOTPMVANDNLAYGPAATSTAGSESSWCACATLTFTSG 120
QY 121 VAGKTWVVGSTGTGDLGSGNOFDIAMPGGGAGIEMGSSQFGGLPGAQYGISRPDCDS 180
DB 121 VAGKTWVVGSTGTGDLGSGNOFDIAMPGGGAGIEMGSSQFGGLPGAQYGISRPDCDS 180
QY 181 FPAPLKPCCQMRPFQONADNPFTTPOVQCPAIVASGCKRNDSSFPVTPPSGG 240
DB 181 FPAPLKPCCQMRPFQONADNPFTTPOVQCPAIVASGCKRNDSSFPVTPPSGG 240
QY 241 GTGTPSTIDAGSGQTSPPGSSGCTSGKMAACCGTGFSGCTTCTTCTTCKKANDYSSQ 298
DB 241 NPSTPT-TPSS-----GGSGCTADKTAQCGSGSGCTNCSGCTCKTINDYHOC 292

RESULT 2

ID 093782 PRELIMINARY; PRT; 305 AA.
AC 093782;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endoglucanase.
GN Egl3.
OS Humicola grisea var. thermoides.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX NCBI_Taxid=5523;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=IF09854;
RX MEDLINE=9144540; PubMed=990729;
RA Takashima S., Ikura H., Nakamura A., Hidaka M., Masaki H., Uozumi T.;
RT "Comparison of gene structures and enzymatic properties between two
RT endoglucanases from Humicola grisea.";
RL J. Biotechnol. 67:85-97(1999).
DR EMBL; AB001107; BA074956.1; -.
DR HSSP; P43316; ZENG.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR009009; Barwin_like.
DR InterPro; IPR00254; CBD_fungal.
DR InterPro; IPR00034; Glyco_hydro_45.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR Prodom; PD001821; CBD_fungal; 1.
DR SMART; SM00236; ECBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
SQ SEQUENCE 305 AA; 32174 MW; 28C979DEBD0771D CRC64;

Query Match 70.5%; Score 1170; DB 3; Length 305;
Best Local Similarity 67.8%; Pred. No. 2,1e-85;
Matches 208; Conservative 44; Mismatches 45; Indels 10; Gaps 4;

QY 1 MRSTPVLRITLAAALPLVSAASGSGSTRYWDCKRSCAMPKAAVSQPYVACDANFOR 60
DB 1 MRSTPVLRITLAAALPLVSAASGSGSTRYWDCKRSCAMPKAAVSQPYVACDANFOR 58
QY 61 LSDENVVSGGNGSAYSCADOTPMVANDNLAYGPAATSTAGSESSWCACATLTFTSG 119
DB 61 LSDENVVSGGNGSAYSCADOTPMVANDNLAYGPAATSTAGSESSWCACATLTFTSG 118
QY 120 PVAGKTWVVGSTGTGDLGSGNOFDIAMPGGGAGIEMGSSQFGGLPGAQYGISRPDCD 179

DB 119 PVAGKTWVVGSTGTGDLGSGNOFDIAMPGGGAGIEMGSSQFGGLPGAQYGISRPDCD 178
QY 180 SEPAPLKPCCQMRPFQONADNPFTTPOVQCPAIVASGCKRNDSSFPVTPPSGG 238
DB 179 SEPAPLKPCCQMRPFQONADNPFTTPOVQCPAIVASGCKRNDSSFPVTPPSGG 238
QY 239 NGGTGTPSTIDAGSGQTS-----PGSSGCTSGKMAACCGTGFSGCTTCTTCKKANDYSSQ 292
DB 239 NGGTGTPSTIDAGSGQTS-----PGSSGCTSGKMAACCGTGFSGCTTCTTCKKANDYSSQ 292
QY 293 DYSYQCL 299
DB 299 DYSYQCL 305

RESULT 3

ID 0810K8 PRELIMINARY; PRT; 235 AA.
AC 0810K8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cellulase precursor (EC 3.2.1.4).
GN CEL45A.
OS Melanocarpus albomyces.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariales incertae sedis;
OC Melanocarpus.
OX NCBI_Taxid=204285;
RN [1]
RS SEQUENCE FROM N.A.
RA Haakana H., Miettinen-Oinonen A., Joutsio J. V., Mantyla A.,
RA Suominen P., Vehmänpää J.;
RT "Cellulase genes from Melanocarpus albomyces: Cloning and High-level
RT Expression in Trichoderma reesei.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ515703; CAD5665.1; -.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0016798; F:hydrolyase activity; acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR009009; Barwin_like.
DR InterPro; IPR00034; Glyco_hydro_45.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
KM Signal; Hydrolyase; Glycosidase.
FT SIGNAL 21
FT CHAIN 22 235
SQ SEQUENCE 235 AA; 25001 MW; BF9850B8366CD76 CRC64;

Query Match 62.6%; Score 1039; DB 3; Length 235;
Best Local Similarity 76.3%; Pred. No. 4.2e-75;
Matches 180; Conservative 27; Mismatches 27; Indels 2; Gaps 1;

QY 1 MRSTPVLRITLAAALPLVSAASGSGSTRYWDCKRSCAMPKAAVSQPYVACDANFOR 60
DB 1 MRSTPVLRITLAAALPLVSAASGSGSTRYWDCKRSCAMPKAAVSQPYVACDANFOR 58
QY 61 LSDENVVSGGNGSAYSCADOTPMVANDNLAYGPAATSTAGSESSWCACATLTFTSG 120
DB 61 LSDENVVSGGNGSAYSCADOTPMVANDNLAYGPAATSTAGSESSWCACATLTFTSG 118
QY 121 VAGKTWVVGSTGTGDLGSGNOFDIAMPGGGAGIEMGSSQFGGLPGAQYGISRPDCDS 180
DB 121 VAGKTWVVGSTGTGDLGSGNOFDIAMPGGGAGIEMGSSQFGGLPGAQYGISRPDCDS 178
QY 181 FPAPLKPCCQMRPFQONADNPFTTPOVQCPAIVASGCKRNDSSFPVTPPS 236
DB 181 FPAPLKPCCQMRPFQONADNPFTTPOVQCPAIVASGCKRNDSSFPVTPPS 234

RESULT 4

Q80110


```

QY 18 VASAAGSGSTRWYDCKPSCAMPKAAVOPVYACDANFQRLSDPNVQSGGSAVS 77
DB 148 VSGASGNETRYDCKPSGWPAGADVTSPVGSCKKGLADNNTQNGCVGSSSTY 207
QY 78 CADQTPAVNDNLAYGPAATSIAGSSSSMCACALFTFSGPVAKTWWQSTGTGDL 137
DB 208 CNDQTPWYVSDIAYGFAAASISGSSBATWCCACELTFTSTAVKAKMVTQVNTGSDL 267
QY 138 GSN---OFDIAMPGGGAGIENGCSQFGGLP---GAQYGGISRPDCDSFPAPLPGCCW 191
DB 268 GSNNGAHEFLDMPGGGAGIENGCAATGCAPTGMCARYGVSASPSKLSALQACCKK 327
QY 192 RFDMPKADNPTFTQVQCPARIVARSCKR 223
DB 328 RFGWPKAANPTMTKYQVTCRKALTKSGSCSR 359

RESULT 7
QY 09JH92 PRELIMINARY; PRT; 219 AA.
AC 09JH92;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
OX NCBI_TaxID=42452;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RT protists in the hindgut of termite Reticulitermes speratus."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045171; BAA98041.1; -.
DR HSSP; P43316; ZENG.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000309; Barwin like.
DR InterPro; IPR000334; Glyco_hydro_45.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
SQ SEQUENCE 219 AA; 23001 MW; 5F2EB81A6BE926CE CRC64;

Query Match 43.6%; Score 724; DB 14; Length 219;
Best Local Similarity 60.3%; Pred. No. 4,9e-50;
Matches 132; Conservative 26; Mismatches 53; Indels 8; Gaps 4;

QY 11 LAALPLVSAASGSGSTRWYDCKPSCAMPKAAVOPVYAC--DANFQRLSDPNVQSG 68
DB 2 LVFTLALISVPGSDGRTTRWDCCKASCAWKKAAVATPVTGKDGTTTVAANDTVKS 61
QY 69 GNGGSAVSCADQTPAVNDNLAYGPAATSIAGSSSSMCACALFTFSGPVAKTWW 128
DB 62 ACDSGEGWYDQAPNVAWNSVANGFAAACCG-ESGACNCCYELTFTSGPVAKTWW 120
QY 139 GSTGTGDLGNSQFDIAMPGGGAGIENGCSQFGGLP---GAQYGGISRPDCDSFPAP 184
DB 121 QVNTNGDLGNSQFDIAMPGGGAGIENGCTQ--SGAPADGWSRGGVSSRSBSQULPSG 179
QY 185 LKPGCCMRPDMFQNDNPTFTQVQCPARIVARSCKR 223
DB 180 LQAGCCMRPDMFQNDNPTFTQVQCPARIVARSCKR 218

RESULT 8
QY 09JH84 PRELIMINARY; PRT; 220 AA.
AC 09JH84;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

```

```

DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
OX NCBI_TaxID=42452;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RT protists in the hindgut of termite Reticulitermes speratus."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045179; BAA98049.1; -.
DR HSSP; P43316; ZENG.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000309; Barwin like.
DR InterPro; IPR000334; Glyco_hydro_45.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
SQ SEQUENCE 220 AA; 23274 MW; 26AF557512EA061 CRC64;

Query Match 43.6%; Score 724; DB 14; Length 220;
Best Local Similarity 59.8%; Pred. No. 4,9e-50;
Matches 131; Conservative 30; Mismatches 50; Indels 8; Gaps 4;

QY 11 LAALPLVSAASGSGSTRWYDCKPSCAMPKAAVOPVYAC--DANFQRLSDPNVQSG 68
DB 3 LKPSCLTISWVGDGRTTRWDCCKASCAWKKAAVATPVTGKDGTTTVAANDTVKS 62
QY 69 GNGGSAVSCADQTPAVNDNLAYGPAATSIAGSSSSMCACALFTFSGPVAKTWW 128
DB 63 SCDDGDXGMYCDQAPNVAWNSVANGFAAACCG-ETGACNCCYELTFTSGPVAKTWW 121
QY 129 GSTGTGDLGNSQFDIAMPGGGAGIENGCSQFGGLP---GAQYGGISRPDCDSFPAP 184
DB 122 QVNTNGDLGNSQFDIAMPGGGAGIENGCTQ--SGAPADGWSRGGVSSRSBSQULPSG 180
QY 185 LKPGCCMRPDMFQNDNPTFTQVQCPARIVARSCKR 223
DB 181 LQAGCCMRPDMFQNDNPTFTQVQCPARIVARSCKR 219

RESULT 9
QY 09JH91 PRELIMINARY; PRT; 219 AA.
AC 09JH91;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
OX NCBI_TaxID=42452;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RT protists in the hindgut of termite Reticulitermes speratus."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045172; BAA98042.1; -.
DR HSSP; P43316; ZENG.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000309; Barwin like.
DR InterPro; IPR000334; Glyco_hydro_45.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
SQ SEQUENCE 219 AA; 23126 MW; A712B83E3CAB041C CRC64;

Query Match 42.7%; Score 709; DB 14; Length 219;
Best Local Similarity 59.4%; Pred. No. 7,7e-49;
Matches 130; Conservative 28; Mismatches 53; Indels 8; Gaps 4;

QY 11 LAALPLVSAASGSGSTRWYDCKPSCAMPKAAVOPVYAC--DANFQRLSDPNVQSG 68

```

```

Db      2  LFTICLISLISGERTIRYMDCKKASCEKKAAYTPVTCAGKGTIRVANSNTYKS 61
Qy      69  GNGGSAYS CADQTPMAVNDNLAYGPAATSIAGSSSSWCCACVALFTSGPVAGKTMV 128
Db      62  CCGGEGYMCYDQAPMAVNDNSVAYGFAAACCAG-ESGACNCYELFTSGPVNGKMTV 120
Qy      129  QSTNGSDLSNQPDLAMPGGGVIENGCSQPGULP----GAQYGGISSEDDCSFPAP 184
Db      121  QVTNNGSDLSNQPDLAIPGGGVIENGCTQ-SGAPSDGWSRYGVSSRSQSQPSG 179
Qy      185  LKPGQMRPDMFQNAUNPTFTFOVQCPAIVASGCKR 223
Db      180  LQAGCMTFDMFQNAUNPS-MRQVTCPSSELTAKTCRN 218

```

```

RESULT 10
QJH83  ID 09JH83  PRELIMINARY; PRT; 220 AA.
AC 09JH83:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
CX NCBI_TaxID=42452;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RT protists in the hindgut of termite Reticulitermes speratus.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
Db      EMBL; AB045167; BAA98037.1; -.
DR HSSP; P43316; ZENG.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR009009; Barwin like.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
SQ SEQUENCE 220 AA; 23108 MW; 97738D831BCFA5F3 CRC64;

```

```

Query Match      42.6%; Score 706.5; DB 14; Length 220;
Best Local Similarity 57.3%; Pred. No. 1.2e-48;
Matches 128; Conservative 32; Mismatches 52; Indels 9; Gaps 5;
Qy      11  LAAALPLVASAASG-SGQSTRYMDCKKSCAMPKAAVSQPYAC--DANFQRLSDPNVQ 67
Db      2  LVEFVSLLASVLFSDSEKTRRYMDCKKSCMEKADVSKPLTTCADGTTIRVANSNDYVK 61
Qy      68  SGNNGSAYSCADQTPMAVNDNLAYGPAATSIAGSSSSWCCACVALFTSGPVAGKTMV 127
Db      62  SGDDGDDGMYCYDQTPMGVNDSYALGFPAALISGG-EKACNCYELFTSGPVNGKMTV 120
Qy      128  VQSTGSDLSNQPDLAMPGGGVIENGCSQPGULP----GAQYGGISSEDDCSFPAP 183
Db      121  VQVTNNGSDLSNQPDLAIPGGGVIENGCTQ-SGAPSDGWSRYGVSSRSQSQPSG 179
Qy      184  PLKPGQMRPDMFQNAUNPTFTFOVQCPAIVASGCKRN 224
Db      180  GLQAGCMTFDMFQNAUNPS-MRQVTCPSSELTAKTCRN 220

```

```

RESULT 11
QJH95  ID 09JH95  PRELIMINARY; PRT; 217 AA.
AC 09JH95:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Family 45 cellulase homologue.

```

```

OS unclassified eukaryotes.
OC Eukaryota.
CX NCBI_TaxID=42452;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RT protists in the hindgut of termite Reticulitermes speratus.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
Db      EMBL; AB045168; BAA98038.1; -.
DR HSSP; P43316; ZENG.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR009009; Barwin like.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
SQ SEQUENCE 217 AA; 22736 MW; 66DD01346B3769DB CRC64;

```

```

Query Match      42.5%; Score 705; DB 14; Length 217;
Best Local Similarity 58.7%; Pred. No. 1.6e-48;
Matches 125; Conservative 33; Mismatches 47; Indels 8; Gaps 4;
Qy      13  VASAASGSGSTRYMDCKKSCAMPKAAVSQPYAC--DANFQRLSDPNVQSGNGGSA 75
Db      7  LASAFDSGKTRRYMDCKKSCMEKADVSKPLTTCADGTTIRVANSNDYVSGDDGDDG 66
Qy      76  YSCADQTPMAVNDNLAYGPAATSIAGSSSSWCCACVALFTSGPVAGKTMVQSTGSG 135
Db      67  FMCYDQTPMGVNDSYALGFPAALISGG-EKACNCYELFTSGPVNGKMTVQVTNNG 125
Qy      136  DLGNSQDLAMPGGGVIENGCSQPGULP----GAQYGGISSEDDCSFPAPLPGQGM 191
Db      126  DLGNSQDLAIPGGGVIENGCTQ-SGAPSDGWSRYGVSSRSQSQPSGLQAGCQM 184
Qy      192  RFDMPQNAUNPTFTFOVQCPAIVASGCKRN 224
Db      185  RFDMPQNAUNPS-MRQVTCPSSELTAKTCRN 217

```

```

RESULT 12
QJH86  ID 09JH86  PRELIMINARY; PRT; 219 AA.
AC 09JH86:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
CX NCBI_TaxID=42452;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RT protists in the hindgut of termite Reticulitermes speratus.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
Db      EMBL; AB045177; BAA98047.1; -.
DR HSSP; P43316; ZENG.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR009009; Barwin like.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
SQ SEQUENCE 219 AA; 23158 MW; ECD686ABEBEJ1D1 CRC64;

```

```

Query Match      42.4%; Score 703; DB 14; Length 219;
Best Local Similarity 58.1%; Pred. No. 2.3e-48;
Matches 125; Conservative 31; Mismatches 51; Indels 8; Gaps 4;
Qy      15  LPLVASAASGSGSTRYMDCKKSCAMPKAAVSQPYAC--DANFQRLSDPNVQSGNG 72

```

Db 6 LTFVGLSLAESGKTRRYWDCCKSGCEKKAIVDKPIDTCAGKDTTRVANSMDTVKSGC6G 65
 QY 73 GSAYSCADQTPWAVNDNLAYGFATSIAGSSSSSWCCACVALLFTPSGPAKTMVQGSTS 132
 Db 66 GEGWICYDQTPWVSVDSTSYGFAAACCCGG-ESGACGCGVDTLFTPSGPAKTMVQITN 124
 QY 133 TGGDLGNSQFDIAMPGGGAGVGFNGCSSQFGGLP-----GAQYGGISRPQCCSPAPLAKG 188
 Db 125 TGGDLGNSQFDIATPGGAGVGIYNGCTAQ-SGAPSDGMSRYSYGVSRSECCQLPSGLQAG 183
 QY 189 CQWRFDPQNDNPTFTFQVQCCPAETVARSQCR 223
 Db 184 CQWRFDPQNDNPTFTFQVQCCPAETVARSQCR 218

RESULT 13

QY0H87 ID QY0H87 PRELIMINARY; PRT; 219 AA.
 AC QY0H87
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Family 45 cellulase homologue.
 OS unclassified eukaryotes.
 OC Eukaryota.
 NCBI_TaxID=42452;
 [1]
 RP SEQUENCE FROM N.A.
 RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
 RT "Diverse genes of family 45 cellulase homologues of the symbiotic
 RT prokists in the hindgut of termite Reticulitermes speratus";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB045176; BAA98046.1; -
 DR HSSP: P43316; ZENG.
 DR GO: GO:0008810; F:cellulase activity; IEA.
 DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro: IPR009009; Barwin like.
 DR InterPro: IPR00334; Glyco_hydro_45.
 DR Pfam: PF02015; Glyco_hydro_45; 1.
 DR PROSITE: PS01140; GLYCOSYL_HYDROL_F45; 1.
 SQ SEQUENCE 219 AA; 23134 MW; 4BDF4EC9ACCT72D CRC64;

Query Match 42.2%; Score 730; DB 14; Length 219;
 Best Local Similarity 57.7%; Pred. No. 4e-48;
 Matches 124; Conservative 34; Mismatches 49; Indels 8; Gaps 4;

QY 15 LPIVSAASGSGSTRYWDCCPSGKAPKAAVQPYAC--DANFQLSPNVQSGNG 72
 Db 6 LPIVSAASGSGSTRYWDCCPSGKAPKAAVQPYAC--DANFQLSPNVQSGNG 65
 QY 73 GSAYSCADQTPWAVNDNLAYGFATSIAGSSSSSWCCACVALLFTPSGPAKTMVQGSTS 132
 Db 66 GEGWICYDQTPWVSVDSTSYGFAAACCCGG-ESGACGCGVDTLFTPSGPAKTMVQITN 124
 QY 133 TGGDLGNSQFDIAMPGGGAGVGFNGCSSQFGGLP-----GAQYGGISRPQCCSPAPLAKG 188
 Db 125 TGGDLGNSQFDIATPGGAGVGIYNGCTAQ-SGAPSDGMSRYSYGVSRSECCQLPSGLQAG 183
 QY 189 CQWRFDPQNDNPTFTFQVQCCPAETVARSQCR 223
 Db 184 CQWRFDPQNDNPTFTFQVQCCPAETVARSQCR 218

RESULT 14

QY0H87 ID QY0H87 PRELIMINARY; PRT; 227 AA.
 AC QY0H87
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Endoglucanase.
 GN EGL4.
 GN Humicola grisea var. thermoides.

OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
 OK NCBI_TaxID=5528;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IFO9854;
 RX MEDLINE=9144540; PubMed=9990729;
 RA Takahina S., Iikura H., Nakamura A., Hidaka M., Maeaki H., Uozumi T.;
 RT "Comparison of gene structures and enzymatic properties between two
 RT endoglucanases from Humicola grisea";
 RL U. Biotechnol. 67:85-97(1999).
 DR EMBL: AB001108; BAA74957.1; -
 DR HSSP: P43316; ZENG.
 DR GO: GO:0008810; F:cellulase activity; IEA.
 DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro: IPR009009; Barwin like.
 DR InterPro: IPR00334; Glyco_hydro_45.
 DR Pfam: PF02015; Glyco_hydro_45; 1.
 DR PROSITE: PS01140; GLYCOSYL_HYDROL_F45; 1.
 SQ SEQUENCE 227 AA; 24240 MW; 873553B76F5C39E4 CRC64;

Query Match 42.1%; Score 698.5; DB 3; Length 227;
 Best Local Similarity 52.7%; Pred. No. 5.5e-48;
 Matches 137; Conservative 26; Mismatches 46; Indels 51; Gaps 8;

QY 9 TTAAALPIVSAASGSGSTRYWDCCPSGKAPKAAVQPYACDANFQLSD-FVQ 67
 Db 6 TTAAALPIVSAASGSGSTRYWDCCPSGKAPKAAVQPYACDANFQLSD-FVQ 63
 QY 68 GCGNCGSAYSCADQTPWAVNDNLAYGFATSIAGSSSSSWCCACVALLFTPSGPAKTM 126
 Db 64 GCGNCGSAYSCADQTPWAVNDNLAYGFATSIAGSSSSSWCCACVALLFTPSGPAKTM 123
 QY 127 VVQSTSTGDLGNSQFDIAMPGGGAGVGFNGCSSQFGGLP-----GAQYGGISRPQCCSPAPLAKG 183
 Db 124 VVQSTSTGDLGNSQFDIAMPGGGAGVGFNGCSSQFGGLP-----GAQYGGISRPQCCSPAPLAKG 183
 QY 184 PLKPCQWRDPQNDNPTFTFQVQCCPAETVARSQCRDSSFPVTPPSGGNGTG 243
 Db 184 PLKPCQWRDPQNDNPTFTFQVQCCPAETVARSQCRDSSFPVTPPSGGNGTG 204
 QY 244 TPTSTAGSGQTSPPGGSGC 263
 Db 205 --LSLPTGQTM--GRSC 219

RESULT 15

QY0H90 ID QY0H90 PRELIMINARY; PRT; 219 AA.
 AC QY0H90
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Family 45 cellulase homologue.
 OS unclassified eukaryotes.
 OC Eukaryota.
 NCBI_TaxID=42452;
 [1]
 RP SEQUENCE FROM N.A.
 RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
 RT "Diverse genes of family 45 cellulase homologues of the symbiotic
 RT prokists in the hindgut of termite Reticulitermes speratus";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB045173; BAA98043.1; -
 DR HSSP: P43316; ZENG.
 DR GO: GO:0008810; F:cellulase activity; IEA.
 DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro: IPR009009; Barwin like.
 DR InterPro: IPR00334; Glyco_hydro_45.
 DR Pfam: PF02015; Glyco_hydro_45; 1.
 DR PROSITE: PS01140; GLYCOSYL_HYDROL_F45; 1.
 SQ SEQUENCE 219 AA; 23037 MW; 372E01641530A9A CRC64;

Query Match 42.1%; Score 698; DB 14; Length 219;

Matches 122; Conservative 36; Mismatches 49

Matches 122; Conservative 36; Mismatches 49; Indels 8; Gaps 47

15 LPLVASAAGSGGSTRYWDCCCKPSCAMPGKAAVSQPYAC--DANFQRLSDENVQSGCNG 72

6 LTFIGLSLADSGKTRYWDCCCKSGCGWEKANVDKPICTCAKDGTTRVASNDTVKSGCDG 65

73 GSAYSCADQTPWAVNDNLAYGFATSIAGSSESSWCCACYALTFTSGPVAGKTMVQSTS 132

66 GTGFMCYDQTPWQVSDSLSYGFAAACCGG-ESGACCGCYELTFTSGPVGKMMIVQITN 124

133 TGGDLCSNQFDIAMPGGVGIFNGCSSQFGGL?---GAQYGGISSRDQCDSPAPLPKG 188

125 TGGDLGSNQFDLAI PGGVG IYNGCTAQ-SGA?SDGWGSR YGGVSSRSRSCS QLP SGLQAG 183

189 CQWRFDWFQNA DNPTFTFQOVQCPAEI VARS GCKR 223

184 CQWREDFQADNP SINESNVKCPSEIIAKTNCNR 218

job time : 46 secs

Job time : 46 secs